

An age-dependent vocal tract model for males and females based on anatomic measurements

Brad H. Story^{a)}

Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85718, USA

Houri K. Vorperian

Vocal Tract Development Lab, Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue
429, Madison, Wisconsin 53705, USA

Kate Bunton

Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85718, USA

Reid B. Durtschi

Vocal Tract Development Lab, Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue
429, Madison, Wisconsin 53705, USA

(Received 30 September 2017; revised 29 April 2018; accepted 1 May 2018; published online 23 May 2018)

The purpose of this study was to take a first step toward constructing a developmental and sex-specific version of a parametric vocal tract area function model representative of male and female vocal tracts ranging in age from infancy to 12 yrs, as well as adults. Anatomic measurements collected from a large imaging database of male and female children and adults provided the dataset from which length warping and cross-dimension scaling functions were derived, and applied to the adult-based vocal tract model to project it backward along an age continuum. The resulting model was assessed qualitatively by projecting hypothetical vocal tract shapes onto midsagittal images from the cohort of children, and quantitatively by comparison of formant frequencies produced by the model to those reported in the literature. An additional validation of modeled vocal tract shapes was made possible by comparison to cross-sectional area measurements obtained for children and adults using acoustic pharyngometry. This initial attempt to generate a sex-specific developmental vocal tract model paves a path to study the relation of vocal tract dimensions to documented prepubertal acoustic differences. © 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5038264>

[ZZ]

Pages: 3079–3102

I. INTRODUCTION

For purposes of studying acoustic characteristics of speech production, and for generating simulated speech sounds, the configuration of the human vocal tract can be represented by the variation in cross-sectional area along a curvilinear axis extending from glottis to lips. Simplifying complex vocal tract shapes as area functions allows for straightforward calculations of resonance frequencies and bandwidths, as well as sound generation based on one-dimensional wave propagation algorithms. Over the past few decades, sets of area functions corresponding to vowels and consonants produced by specific talkers have been obtained by utilizing various three-dimensional imaging modalities and image analysis techniques (e.g., Baer *et al.*, 1991; Yang and Kasuya, 1994; Narayanan *et al.*, 1995, 1997a, 1997b; Alwan *et al.*, 1997; Story *et al.*, 1996; Story and Titze, 1998; Story *et al.*, 2001; Story, 2005b; Takemoto *et al.*, 2006). While these data sets are valuable contributions to understanding various aspects of speech production, they are all based on the vocal tract configurations of *adult* talkers, and hence represent adult speech. There is a paucity of similar data representing vocal tract shapes produced by children as

they grow and develop from infancy through puberty. Consequently, it is difficult to simulate child-like speech as well as investigate various developmental acoustic phenomena that remain unexplained. For example, formant frequencies change very little during the first two years of life (Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982) even though increases in vocal tract length (VTL) suggest they should decrease (Vorperian *et al.*, 2005, 2009). There are also reported sex differences in formant frequencies by age four (Perry *et al.*, 2001), despite there being essentially no sex differences in VTL (Fitch and Giedd, 1999; Vorperian *et al.*, 2005, 2009). Questions such as these need to be addressed with knowledge of how the vocal tract is configured during production of speech sounds.

With the exception of area functions reported by Yang and Kasuya (1994) for three vowels /i, a, u/ produced by an 11 yr old boy, there are, to the authors' knowledge, no reported sets of area functions of vowels or consonants measured for children at any age prior to puberty. There is good reason for this absence of area function data: Collection of volumetric image sets requires the talker to perform static speech sounds repeatedly, typically in supine position, in a medical diagnostic environment that can be intimidating, sometimes even to adults. Thus, collecting image sets of this type based on speech sound production from children would be, in most cases, impractical.

^{a)}Electronic mail: bstory@email.arizona.edu

Medical imaging, such as x ray, computed tomography (CT), and magnetic resonance imaging (MRI), however, has been used to secure anatomic data on the size and shape of vocal tract structures by either teaming up with other prospective imaging studies, such as the anatomic data from MRI reported by [Fitch and Giedd \(1999\)](#), or by retrospectively collecting scans of infants, children, and adolescents who were imaged (many subjects representing a large age range) for various medical reasons using strict criteria to exclude cases or diagnoses that alter or affect the clear visualization of the anatomic structures that contribute to speech production ([Vorperian et al., 1999, 2005, 2009, 2011](#); [Barbier et al., 2015](#)). Although indispensable for understanding the structural development of the speech production mechanism, these data do not provide an indication of the actual vocal tract shapes that children configure to produce speech sounds at successive stages of development.

A typical approach to studying the relation of vocal tract shape to acoustic characteristics in children is to generate *hypothetical* area functions using some type of model based on knowledge, however limited, of the anatomic structures at a given age (cf. [Lieberman and Crelin, 1971](#); [Lieberman et al., 1972](#); [Nordström, 1977](#); [Goldstein, 1980](#); [Boë, 1999](#); [Boë et al., 2007](#); [Ménard et al., 2004, 2009](#)). An early attempt was reported by [Lieberman and Crelin \(1971\)](#) who measured cross-sectional areas of plaster casts made from newborn infant and adult cadavers. Assuming these to approximate neutral area functions (i.e., roughly an /ə/), they perturbed them into possible configurations of the vowels /i/, /a/, and /u/. They concluded that, because of the relatively short pharyngeal section of the infant vocal tract, infants would not be able to produce formant frequencies that present the same degree of acoustic contrast between vowels as adults. The suggestion was that a 1:1 ratio of the lengths of the oral cavity to the pharyngeal cavity is necessary to allow the vocal tract to take on “two-tube” configurations that produce distinct formant frequencies characterizing the corner vowels in adult speech, whereas this ratio has been shown to be roughly 1.5:1 in infants. In more recent publications regarding infant, Neanderthal, and nonhuman primate vocalizations, [Lieberman \(2006, 2007, 2012\)](#) has maintained that the ratio of the horizontal and vertical aspects of the vocal tract is a primary factor in the emergence and development of human speech production abilities. Indeed, based on extensive analysis of longitudinally-collected midsagittal radiographs of children engaged in nonspeech tasks, [Lieberman and McCarthy \(1999\)](#) and [Lieberman et al. \(2001\)](#) showed that a 1:1 ratio is not achieved until about 6–8 yrs of age. Similar results were reported by [Fitch and Giedd \(1999\)](#) and [Vorperian et al. \(2005\)](#).

A different approach was reported by [Nordström \(1977\)](#) who generated hypothetical area functions of infant and adult female vocal tracts by modifying measured adult male area functions from [Fant \(1960\)](#). Modifications consisted of reducing the pharyngeal cavity size, both in length and volume. Although calculations of vocal tract resonance frequencies for the infant and female area functions were in the direction of observed formant values, they did not confirm the formant scaling factors that had been previously proposed by [Fant \(1966, 1975\)](#). Nordström concluded that children and adult females must use vocal tract configurations

that differ in ways that are not captured by simple length and volume scaling adjustments.

Other efforts to generate hypothetical area functions have been based on developing articulatory models for which a given set of anatomically-based parameters dictate the positions of speech articulators such that they form a specific vocal tract shape in the midsagittal plane. An area function is then obtained by transforming midsagittal cross-distances to cross-sectional areas (cf. [Heinz and Stevens, 1965](#); [Beautemps et al., 1995](#)). This was the approach of [Goldstein \(1980\)](#) who compiled growth curves of 19 structures based on a variety of published anatomic data, then used curve fits of the data to generate age-dependent parameter values for an articulatory model (specifically, [Mermelstein, 1973](#)). The result was a model of the vocal tract that could “grow” from infancy to adulthood and produce plausible child-like acoustic characteristics. Although not focused on children’s speech *per se*, [de Boer \(2010\)](#) used a highly simplified version of the [Mermelstein \(1973\)](#) articulatory model to show that a roughly 1:1 oral cavity to pharyngeal ratio produced the largest range of acoustic contrast between vowels, thus supporting the previously discussed hypothesis of Lieberman. [Badin et al. \(2014\)](#) disputed de Boer’s findings because of the absence of a lip section in his formulation of the articulatory model. An independently-controlled lip section, it was argued, would allow for compensation of a wide range of laryngeal heights (and hence pharyngeal cavity lengths) to maintain a large [F_1, F_2] vowel space.

Another series of studies have demonstrated the effects of vocal tract growth on speech by modifying the parameters of the Variable Linear Articulatory Model (VLAM) ([Boë and Maeda, 1998](#); [Ménard et al., 2004](#)). The VLAM is based on a midsagittal representation of the vocal tract originally developed by [Maeda \(1979, 1990\)](#) in which linear combinations of vocal tract shaping components obtained from factor analysis of cineradiographic images of two adult female talkers (native speakers of French) were used to reconstruct vocal tract configurations. Minimally, the parameters of the model control the shape and position of the tongue, lips, jaw, and larynx height, but may also include specification of palate height and pharyngeal wall contour ([Boë et al., 2006](#)). The VLAM operates by first specifying parameter values of the Maeda model in the original adult domain, resulting in a midsagittal vocal tract configuration. To transform to a child-like vocal tract, two scaling factors, one for the pharynx and another for the oral cavity, are used to impose age-dependent and nonuniform modifications of both the length axis as well as the distance across the tract. The values of the two scaling factors across age are based on data and results reported in [Goldstein \(1980\)](#). An area function is obtained by converting the midsagittal cross-distances to areas via the approach given in [Heinz and Stevens \(1965\)](#), and is assumed to be the same for children as adults. Similar in concept to [Nordström \(1977\)](#), the notion is that reduction of the cross-sectional area (or midsagittal cross-distance) is proportionally coincident with contraction of the VTL axis. The VLAM has been used extensively to simulate vocal tract growth, and has demonstrated hypothetical configurations that could perhaps be achieved by children during speech production ([Boë et al., 2013](#)). By necessity, however, their means of

validation has been based primarily on comparison of calculated formant frequencies to those reported in the literature rather than to actual measurements of vocal tract shapes.

Essentially the same model has been implemented by [Oohashi et al. \(2017\)](#) in a system that maps formant frequencies measured from acoustic recordings of children to hypothetical vocal tract shapes. These can be used to show possible differences in articulatory configurations for production of a variety of vowels at specific ages. The Maeda articulatory model was also used by [Callan et al. \(2000\)](#) in a study of motor control strategies as the vocal tract grows and develops. The parameters of the articulatory model were modified with developmental growth curves of various anatomic structures that had been measured from magnetic resonance (MR) images. Coupled with the DIVA neural network model ([Guenther et al., 1998](#)), different strategies were learned by the model for producing vowels at different ages to compensate for developmental morphological changes.

As summarized by [de Boer and Fitch \(2010\)](#), vocal tract models can be roughly categorized as geometric or statistical. A model such as the one described by [Mermelstein \(1973\)](#) is a geometric system of articulatory structures defined by anatomic measurements, whereas a model like [Maeda's \(1979, 1990\)](#) is statistical in the sense that a given vocal tract configuration is generated by combining shaping patterns that have been derived from some type of statistical analysis [e.g., principal component analysis (PCA) or factor analysis] of articulatory data obtained during speech production. Although each type has advantages and disadvantages, the use of statistical models (e.g., [Maeda, 1979, 1990](#)) to investigate articulatory and acoustic possibilities of the developing human and nonhuman primate vocal tract, particularly the VLAM approach, has been criticized on the grounds that it inappropriately reconfigures statistically-derived shaping patterns for an adult vocal tract to fit the anatomical structure of a child (cf. [Lieberman, 2012](#)). In the view of [de Boer and Fitch \(2010\)](#), this approach suffers from “logical circularity” because the key aspects of the adult vocal tract (i.e., the ability to produce large discontinuities in cross-section in the pharynx relative to the oral cavity) are maintained in the reconfiguring process, and thus overestimates the abilities of a child talker. The suggestion is that the anatomy should define and constrain a model of the child’s vocal tract shape, not be used to modify an already mature one.

There is clearly no direct path for modeling the vocal tract shape of children. Advances in magnetic resonance imaging techniques, along with articulography and ultrasound, continue to add to our knowledge of anatomy, tongue configuration, and articulatory movement during speech development, but measurement of vocal tract area functions produced by children during speech production remains elusive. Thus, development of models must rely, in large part, on using documented prepubertal differences of vocal tract anatomy (e.g., [Lieberman and McCarthy, 1999; Fitch and Giedd, 1999; Vorperian et al., 2011](#)) and acoustic data (cf. summary in [Vorperian and Kent, 2007](#)) to inform and validate possible configurations of the vocal tract for children’s production of vowels and consonants. Attempts to build age-dependent and sex-specific vocal tract models are needed in

order to explore the possible acoustic characteristics and perception of children’s speech.

The aim of this study was to take a first step toward constructing a developmental and sex-specific version of the parametric area function model described in [Story \(2005a\)](#). For vowel production, the adult form of this model generates a vocal tract configuration by deforming a neutral tract shape with a linear combination of two shaping patterns whose contributions are controlled by the amplitudes and polarities of two parameters. Informed by anatomic data based on direct measurements from imaging studies of a large cohort of children and adults, the model components were warped along their length axis and scaled in cross-section to become representative of a male or female child at a given age. This approach does indeed consist of transforming a statistically-based adult model to one that is child-like, but the anatomically-based length warping process generates nonuniform, continuous age and sex-dependent scaling functions that modify the cross-dimension of the vocal tract such that the pharyngeal portion is attenuated to a greater degree than the oral cavity. That is, the adult-like abilities to expand and constrict the back and front portions of the vocal tract are considerably altered by processes based on anatomical measurements. In addition, the two model parameters that control vocal tract shape have been shown to maintain a one-to-one relation with the first two acoustic resonances ($[f_{R1}, f_{R2}]$) of the vocal tract, thus the transformation of the model from adult to child will indicate whether the one-to-one mapping is retained, and how the vowel space may be expanded, contracted, or deformed under those conditions. Modeling the area function directly, rather than in the midsagittal plane, is also an advantage because it eliminates the need for converting cross-distances to areas ([Heinz and Stevens, 1965; Mermelstein, 1973](#)), an operation that is essentially unknown for children’s vocal tracts (cf. [McGowan et al., 2012](#)). The resulting model was assessed qualitatively by projecting hypothetical vocal tract shapes onto selected midsagittal images of children in the cohort, and quantitatively by comparison of formant frequencies produced by the model to those reported in the literature. An additional validation of modeled vocal tract shapes was made possible by comparison to cross-sectional area measurements obtained for children and adults using acoustic pharyngometry.

The article is arranged such that the area function model is briefly reviewed in Sec. II. Measurement and analysis of anatomic landmarks are presented in Sec. III, followed by the description of a technique for warping the length axis of the vocal tract, and scaling the cross-dimensions of the model components based on the anatomic data in Sec. IV. Application of the technique is demonstrated in Sec. V along with assessment of the plausibility of hypothetical area functions generated by the model, and comparison of their acoustic resonance frequencies to measured formant values. In most sections, “method” and “results” are combined in order to maintain the sequential nature required for describing the process of transforming the adult model to various stages of child development.

II. AREA FUNCTION MODEL OF THE VOCAL TRACT

The model of the vocal tract is based on a PCA of vowel area functions of an adult male talker ([Story, 2005a, 2009](#)).

The specific set used to build the model for this study were the ten target vowels /i, ɪ, ε, æ, ʌ, ə, ɔ, ɒ, u, ʊ/ reported in Story *et al.* (1996). Prior to performing the PCA, each area function was smoothed with a 3-point moving average filter and the entry area just above the glottis was set to 0.5 cm^2 . These preprocessing steps assured that the PCA represented the primary features of the overall vocal tract shape and suppressed somewhat the idiosyncrasies of the original talker. The processed area functions were then converted to equivalent diameters (see Story, 2005a, 2009) and the PCA was performed. The vocal tract area function can be specified as

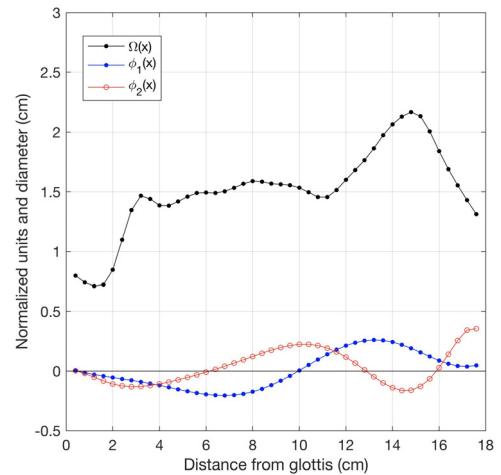
$$A(x) = \frac{\pi}{4} [\Omega(x) + q_1\phi_1(x) + q_2\phi_2(x)]^2, \quad (1)$$

where x is the distance from the glottis and $\Omega(x)$, $\phi_1(x)$, and $\phi_2(x)$ are the mean (neutral) vocal tract diameter function and the first two principal components [henceforth referred to as “modes” as in Story (2005a)], respectively. These three functions, along with the vowel-dependent coefficients, q_1 and q_2 , are provided in numerical form in the Appendix. For this study, all components of the model and resulting area functions consisted of $N_x = 44$ contiguous area elements ordered from the glottal to lip end of the vocal tract. For the adult male version, each element has an assumed length of $L(i) = 0.4 \text{ cm}$, where i is the element number. The value of x corresponding to the i th element is then

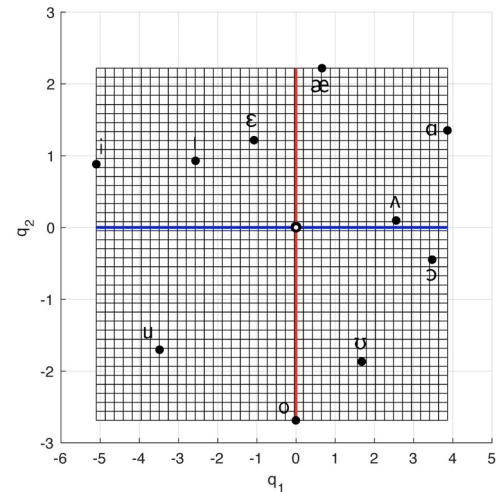
$$x(i) = \sum_{z=1}^i L(z), \quad i = [1, N_x], \quad (2)$$

and results, in this case, in an overall VTL of $x(N_x) = 17.6 \text{ cm}$. Because $L(i)$ was held constant for all elements, any area function generated with Eq. (1) will have the same VTL, regardless of vowel target. Although this approximation underestimates the VTL for some vowels and overestimates it for others, the segment lengths and VTLs obtained from the imaging dataset described in Sec. III are based on a non-speech condition of normal respiration and hence do not provide an indication of vowel-dependent length change. Thus, at this stage of model development a constant VTL was maintained for all vowels.

The $\Omega(x)$, $\phi_1(x)$, and $\phi_2(x)$ functions are shown in Fig. 1(a). The $[q_1, q_2]$ coefficient pairs that reconstruct the original ten area functions are indicated as black dots in Fig. 1(b), whereas the mesh represents 1600 $[q_1, q_2]$ pairs that span the range of each coefficient dimension. Each pair was used to construct a unique area function from which resonance frequencies (formants) were determined by calculating a frequency response function based on a transmission line approach (Sondhi and Schroeter, 1987; Story *et al.*, 2000). This calculation included energy losses due to yielding walls, viscosity, heat conduction, and acoustic radiation at the lips; side branches such as the piri-form sinuses were not considered for this study. The resonance frequencies (f_{R1}, f_{R2}, \dots)¹ were determined by finding the peaks in the frequency response functions with a peak-picking algorithm (Titze *et al.*, 1987). The result is the mesh shown in Fig. 1(c) that represents the acoustic vowel space, in terms of the $[f_{R1}, f_{R2}]$ resonances, produced by deformations of the two modes superimposed on the neutral vocal tract shape.



(a)



(b)

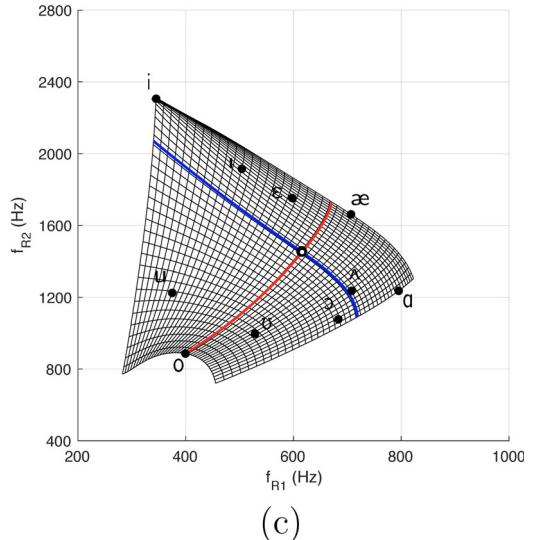


FIG. 1. (Color online) Demonstration of the adult-based area function model. (a) Neutral diameter function and two modes (i.e., principal components), (b) $[q_1, q_2]$ grid spanning the range of each coefficient; and (c) $[f_{R1}, f_{R2}]$ resonance frequency grid corresponding to the coefficient grid. The black dots in (b) and (c) indicate the coefficient pairs and corresponding resonance frequencies associated with the original ten area functions, respectively.

Sections III and IV describe a process by which the components of this model were systematically transformed, based on anatomic data, into a hypothetical vocal tract representative of male and female children across an age range from birth to 12 yrs, as well as to an adult female.

III. ANATOMIC MEASUREMENTS OF THE VOCAL TRACT

Transformation of the area function model from adult to child was accomplished with developmental anatomic data. Using an extant medical imaging database that was retrospectively procured by the Vocal Tract Development Laboratory following University of Wisconsin-Madison's Institutional Review Board approval, 132 CT (i.e., x ray computed tomography) studies of typically developing individuals were randomly selected for the purpose of measuring segment lengths along the vocal tract from the midsagittal plane. The pediatric cohort of this dataset consisted of 112 (58 M; 54 F) CT cases between the ages of birth to 12 yrs, and the adult cohort consisted of 20 (10 M; 10 F) CT cases ages 18 to 25 yrs. In all cases, the CT studies were done for various medical conditions that did not affect the growth of head and neck structures and were classified to have a Class I bite. All studies were acquired using General Electric helical CT scanners (General Electric Healthcare, Chicago, IL), and the images were stored in Digital Imaging and Communications in Medicine format. All images were visually inspected to ensure that the entire vocal tract could be visualized. Details on image acquisition parameters is provided in [Vorperian et al. \(2009\)](#) and [Kelly et al. \(2017\)](#).

A. Location of points along the vocal tract centerline

As shown in Fig. 2, image-based measurements were facilitated by initially constructing a centerline curve that traverses the midpoints between the inferior (soft tissue) and superior (hard and soft tissue) boundaries of the vocal tract wall, extending from the level of the true vocal folds (point A) to the intersection of a line drawn tangentially to the anterior aspect of the upper (maxillary) and lower (mandibular) lips at the level of the stomion (most anterior point of contact between the upper and lower lips; point E). The centerline, as well as landmark points along the centerline were placed independently by two trained researchers and discrepancies resolved with guidance from a head and neck radiologist prior to finalizing landmark placement for measurements. For additional methodological detail, see [Vorperian et al. \(2009\)](#). The curvilinear distance along the centerline, from point A to E, is the overall VTL. Because the analyzed images were not collected during production of speech sounds, information is not available regarding possible VTL differences across vowels.

Each centerline was extracted from the midsagittal image as a matrix of (x, y) pairs and, regardless of actual length, was subsequently resampled to contain 44 rows. The resampling was performed for the sake of efficiency in relating the measurements to the components of the area function model. Allowing point A to serve as the origin, three reference points, B, C, and D, indicating the location of selected anatomic reference landmarks along the centerline were then determined according to the following specifications:

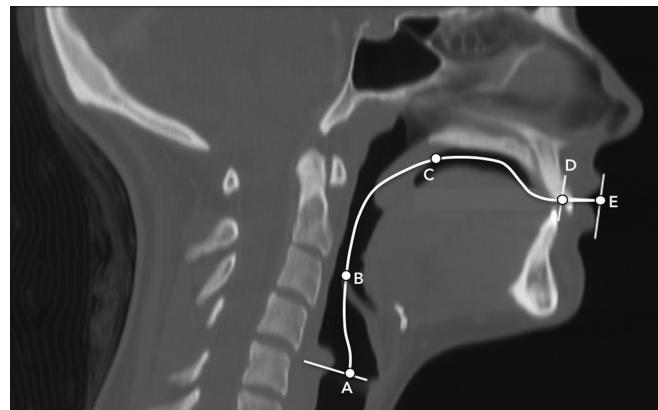


FIG. 2. Illustration of anatomic landmarks measured along the vocal tract centerline profile. A = glottis, B = superior aspect of epiglottis, C = junction of the HPSP, D = posterior aspect of the upper and lower lips at the level of the stomion, E = anterior aspect of the upper and lower lips at the level of the stomion.

Point B: The point along the centerline that corresponded to the superior extent of the epiglottis. Since the epiglottis is a leaf shaped structure, parasagittal slices were examined to ensure that its tip was accurately presented in the midsagittal plane.

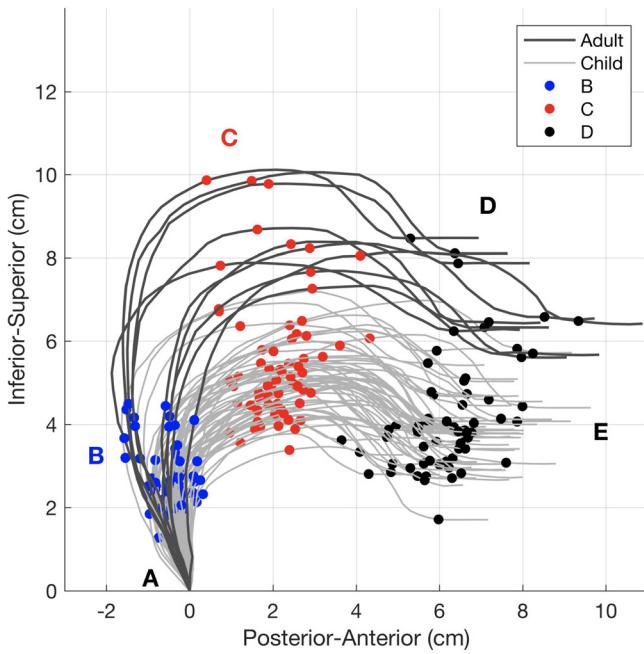
Point C: The point located at the intersection of the vocal tract centerline with the hard palate/soft palate reference line, an oblique line placed at the beginning of the hard palate/soft palate overlap, as guided by the nasal septum, and extending to the superior border of the anterior tongue. This reference point represents the junction of the hard palate and soft palate (HPSP).

Point D: The point located at the intersection of the vocal tract centerline and the posterior or buccal aspect of the upper and lower lips. This point is an approximate indicator of the location of the anterior aspect of the teeth/incisors relative to the posterior aspect of the lips.

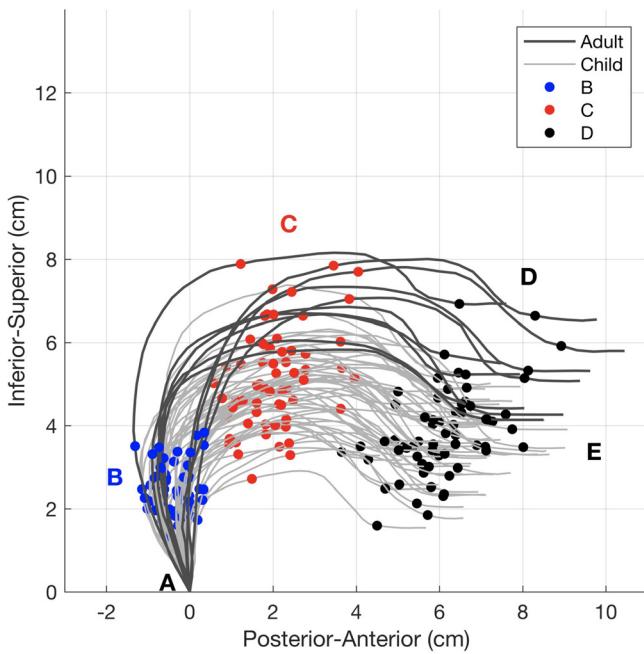
Vocal tract centerlines for all 132 cases are shown in Fig. 3, with the males in the upper panel and females in the lower panel. The origin for each profile is landmark A (glottis), and the end is landmark E (lips). The pediatric cases are plotted with gray lines, and the adults are shown with black lines. The measured points B, C, and D are indicated with dots placed on each centerline, and the glottis and lips are assumed to be points A and E. Although there is observable variation in both the overall shape of the vocal tract centerline and the location of each measured point, it is clear that the adult male centerlines demonstrate a larger vertical extent than either children or adult females, a well-documented aspect of the adult male vocal tract.

B. Analysis of vocal tract segment lengths

Based on the five points (A–E) located along each vocal tract profile, curvilinear distances of segments AB, BC, CD, and DE were measured. This was accomplished by accumulating the Euclidean distances between successive sample points along the profile, within each segment. The measured segments represent approximately the lower pharynx, upper pharynx, oral cavity, and lip region. Separated by sex, data points for each segment were averaged within 14 age bins defined



(a) Male



(b) Female

FIG. 3. (Color online) Collection of vocal tract centerline profiles of a pediatric cohort of 58 males and 54 females between the ages of birth to 12 yrs, and an adult cohort consisting of 10 male and 10 female cases ages 18 to 25 yrs. The solid points plotted on each profile indicate the locations of the anatomic landmark measurements. (a) male and (b) female.

as 0–0.5 yrs, 0.5–1.5 yrs, 1.5–2.5 yrs,..., 11.5–12.5 yrs, and adult. The means and standard deviation (σ) for the male and female data are shown in Figs. 4(a) and 4(b), respectively, as a function of age where year 0 represents the first bin of 0–0.5 yrs, year 1 represents the second bin of 0.5–1.5 yrs, and so forth. The gray zone in each plot is the age range for which there were no data points. The overall VTL for the male and

female pediatric cases ranged from 8–15.4 and 8–15.1 cm, respectively. The adult male VTLs ranged from 15.8–18.8 with a mean value of 17.6 cm ($\sigma = 0.89$ cm), whereas the range was 14–17.5 cm for females with a mean of 15.6 cm ($\sigma = 1.14$ cm).

To parameterize the growth of both segment length and location of the measured points along the vocal tract from 0 yrs to adult, a double logistic equation (cf. [Barbier et al., 2015](#)) was fit to each segment of the male and female averaged data sets. The equation, which specifies segment length L_s as a function of age χ (in years), was of the form $L_s(\chi) = L_1/(1 + \exp^{-[a_1(\chi - x_1)]}) + (L_2 - L_1)/(1 + \exp^{-[a_2(\chi - x_2)]})$ where L_1 and L_2 are amplitude values of the two plateaus in the curve, x_1 and x_2 are age values of the inflection points, and a_1 and a_2 are the slopes at the inflection points. The parameters of the equation were determined for each data set with a nonlinear regression algorithm available in MATLAB (i.e., curve fitting toolbox, [Mathworks, 2016](#)), and then adjusted slightly to ensure a smooth curve. The resulting curve fits are shown as solid lines passing through each set of mean segment lengths in Figs. 4(a) and 4(b).

From Figs. 4(a) and 4(b) it can be observed that segment AB for females follows a nearly linear path between the ages of 0 and 12 yrs, and then plateaus toward the adult segment length. In contrast, the growth of segment AB for males is more rapid in the 0–2 yr range and then steadily increases almost linearly to adulthood; at all ages except at 0 yrs, the length of segment AB is larger for males than females. Segment BC, the length of the upper pharyngeal portion of the vocal tract, grows rapidly between 0 and 4 yrs for males and then exhibits essentially linear growth until the adult length is attained, but for females there is more curvature throughout the entire age range. This results in smaller BC segment lengths for the female from 0–4 yrs, similar lengths from 4–10 yrs, and then smaller values than the male from 10 yrs to adulthood. The growth of segment CD, the oral cavity between the HPSP junction and incisors, is similarly curved for males and females across all ages, showing rapid growth between 0 and 2 yrs. The magnitude of the segment, however, is slightly larger for females up to 7 yrs, at which point the male CD segment length becomes larger. Segment DE is also fairly similar for both males and females up to about 12 yrs, and then increases more for the male vocal tract as growth progresses toward adulthood. Interestingly, the length of BC and CD segments for the male at the earliest age are nearly equal, then the length of segment CD increases more rapidly than the length of BC, and maintains a greater length until their growth curves cross over each other sometime between 12 yrs and adult. The CD segments for the female, however, are longer than the BC segments across the entire age range, but do become nearly equal at the adult stage. The difference in the growth curves of the BC and CD segments reflect the extensive growth in pharyngeal length that takes place during puberty for males, but not females.

The curves representing the four segments in Figs. 4(a) and 4(b) were summed to generate overall male and female VTL curves as a function of age. These are plotted in Fig. 4(c) and indicate that the VTL is nearly equal for males and females at 0 yrs, and then there is a somewhat

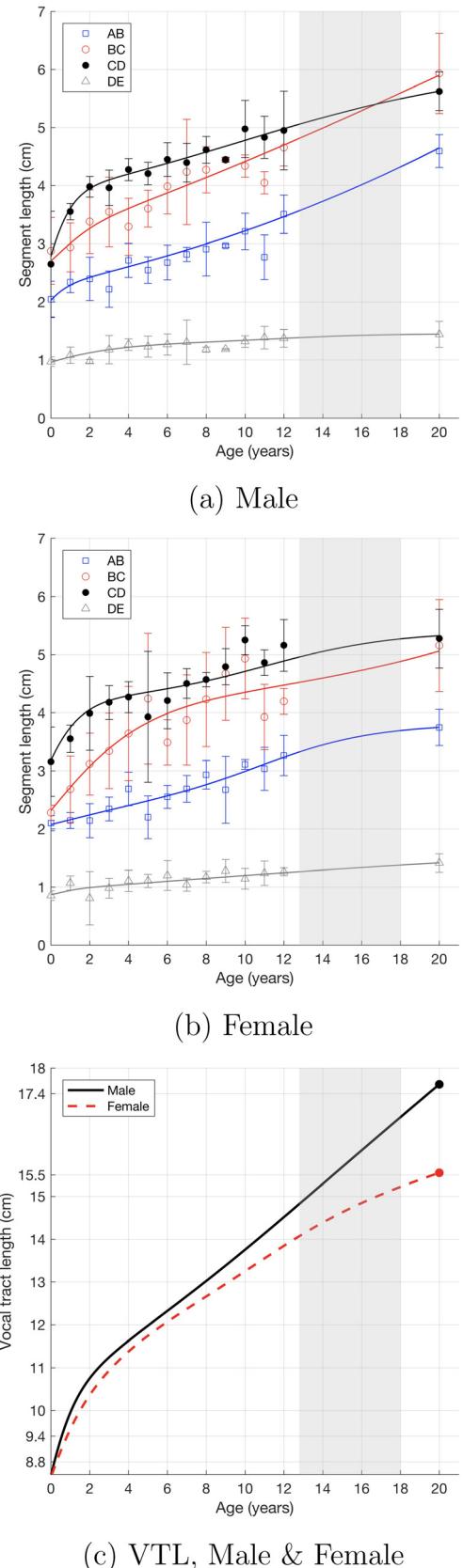


FIG. 4. (Color online) Mean lengths and standard deviations of the four segments (AB, BC, CD, and DE) measured along the vocal tract profile for the pediatric and adult cohorts, shown as a function of age in years: (a) male and (b) female. Double logistic curves fitted to the data are shown as solid lines. (c) Overall VTL for both male (solid line) and female (dashed line) based on the sum of the four segments in the (a) and (b) plots. The gray zone in each panel represents the age range in which there were no data points.

more rapid growth for males between 0 and 3 yrs. From 3–7 yrs the male and female VTL growth follows a similar curve, but the VTL magnitude is slightly larger for males. After 7 yrs, the male and female VTL curves continue to diverge as they move toward adult lengths. The VTL values across age and sex are nearly the same as reported by Vorperian *et al.* (2009) and Fitch and Giedd (1999).

Another way of viewing the curves and the segments they represent is to plot them as functions of age and cumulative tract length as shown in Figs. 5(a) and 5(b). The growth of each segment can be observed from early age at the bottom to adulthood at the top, and overall VTL is indicated by the rightmost curve. The maximum VTL attained for either males or females, based on this data set, is shown by the thin vertical line at the far right of each plot. Overall, after about 4 yrs, the segments can be seen to increase in a slightly more linear fashion across age for males than females. In Figs. 5(c) and 5(d), the length of each segment was normalized to the total VTL at each age. In this view, segments can be seen as a proportion of total tract length, regardless of their absolute length. The thin, vertical lines on each plot indicate the normalized location of the points B, C, and D at the earliest age, so that they can more easily be compared to these same locations at later ages. For either male or female, the largest location shift from childhood to adulthood is observed for point C, the approximate junction of the hard and soft palates, and is indicative of the increased vertical extent of the vocal tract with age.

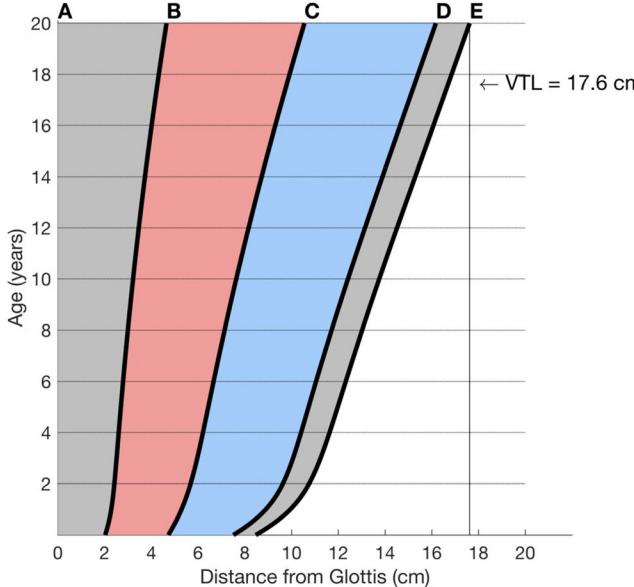
IV. TRANSFORMATION OF THE VOCAL TRACT AREA FUNCTION MODEL FROM ADULT TO CHILD

The segment lengths and normalized double logistic curves shown in Fig. 5 form the basis of the method for warping the length and scaling the cross-sectional dimensions of the vocal tract in order to transform the adult-based model described in Sec. II into an age-dependent child-like version. Warping the length axis consists of adjusting the four segment lengths corresponding to an adult vocal tract such that they become equal to those of a child at a target age. An age-dependent scaling function is then derived from the length warping process that non-uniformly modifies the cross-dimensions of the $\Omega(x)$, $\phi_1(x)$, and $\phi_2(x)$ components of the model.

A. Length axis warping

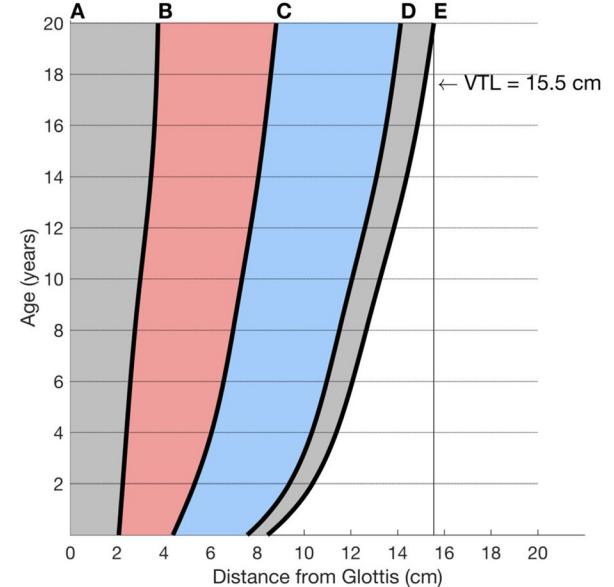
The first step toward warping the length axis is selection of points B, C, and D from the normalized segment curves [Figs. 5(c) and 5(d)] that are deemed representative of the adult vocal tract model. These comprise a vector defined as $\alpha = [0 \ B \ C \ D \ 1]$, where the 0 and 1 represent points A and E in the normalized domain. Since the model in Sec. II is based on an adult male, the vector was taken to be the endpoints (age 20) of the male growth curves in Fig. 5(c). This yields a normalized vector for the adult male vocal tract of $\alpha = [0 \ 0.2639 \ 0.5988 \ 0.9180 \ 1]$, and overall $VTL_\alpha = 17.6$ cm. A second normalized location vector, β , is then selected that represents the target age and sex to which the adult model will be transformed. As an example, a vector selected from

Male

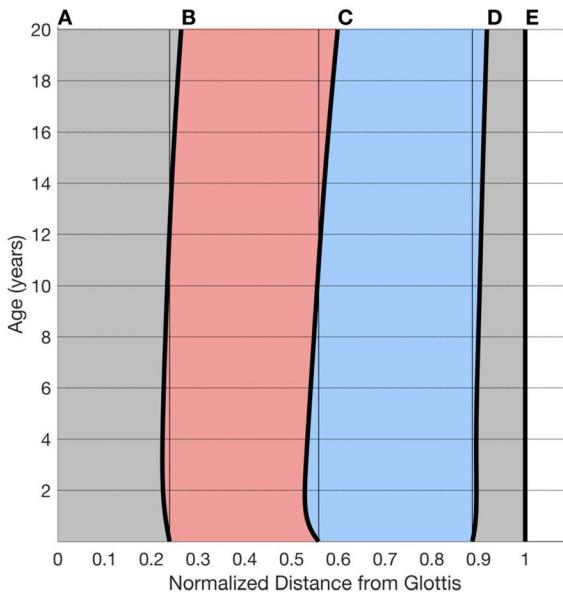


(a) Male, absolute distance

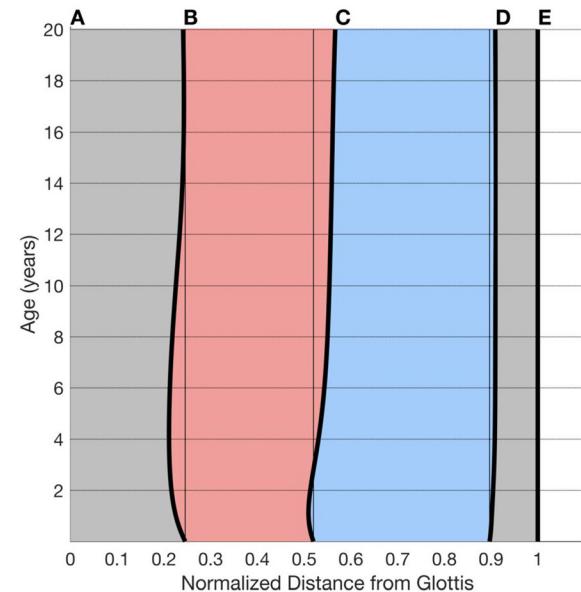
Female



(b) Female, absolute distance



(c) Male, normalized distance



(d) Female, normalized distance

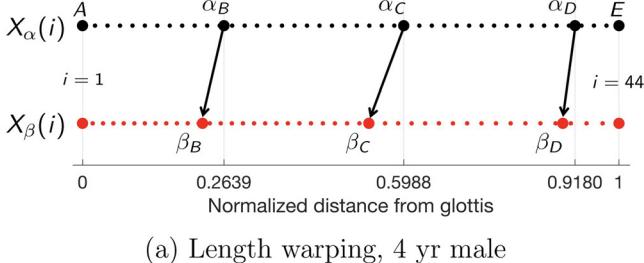
FIG. 5. (Color online) Illustration of the vocal tract segment lengths as a function of age and distance from the glottis, where plots in the left and right columns represent males and females, respectively. The top row of plots shows the segments as absolute length in cm, whereas the bottom row recasts the segments as a function of normalized distance from the glottis. The thin vertical lines in both (c) and (d) indicate the normalized location of the segment boundaries at the earliest age.

the male set of normalized curves [i.e., Fig. 5(c)] at 4 yrs of age would be $\beta = [0 \ 0.2240 \ 0.5341 \ 0.8952 \ 1]$ where the absolute VTL is $VTL_\beta = 11.6$ cm.

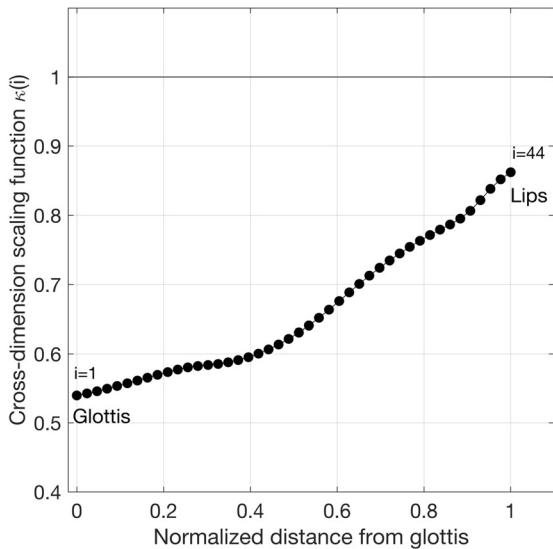
The next step is to generate a normalized length vector, uniformly incremented from glottis to lips, that represents the length axis of adult male vocal tract model as

$$X_\alpha(i) = \sum_{z=1}^i \frac{z-1}{N_x - 1}, \quad i = [1, N_x], \quad (3)$$

where $N_x = 44$ corresponds to the number of elements. This vector is shown schematically in Fig. 6(a) as the upper line of dots, which are equally spaced from 0 (glottis) to 1 (lips). The α vector representing the anatomic locations B, C, and D for the adult male are superimposed on $X_\alpha(i)$. The second line of dots in Fig. 6(a) forms a warped length axis, X_β produced by projecting the points in the α vector to the locations specified in the β vector. This results in a non-uniformly incremented length axis where the distance



(a) Length warping, 4 yr male



(b) Scaling function, 4 yr male

FIG. 6. (Color online) Demonstration of warping the length axis and determining a cross-dimension scaling function for a 4 yr-old male. (a) The upper row of dots is the uniformly spaced length vector X_α where α_B , α_C , and α_D are the anatomic landmarks for the adult male case (the tick marks are located at the locations). The lower row of dots is the warped length vector X_β that results from shifting the α locations in the upper row to those of β_B , β_C , and β_D for a 4 yr-old male. (b) Cross-dimension scaling function $\kappa(i)$ shown as a function of normalized distance from the glottis, ordered from glottis to lips.

between some points has been compressed, and expanded between others. Computationally, this operation was performed by letting β become a function of α , and then resampling X_α with a piecewise cubic interpolation (Fritsch and Carlson, 1980; implemented as the “pchip” algorithm available in MATLAB, Mathworks, 2016) to generate X_β .

B. Cross-dimension scaling

In addition to adjusting the length axis of the vocal tract model to be child-like, the cross-dimension from glottis to lips must also be modified to account for the smaller space available to produce expansions and constrictions when the dimensions of the vocal tract are reduced. Thus, transforming the adult-based model of Sec. II requires a scaling function that attenuates the $\Omega(x)$ function and the modes $\phi_1(x)$ and $\phi_2(x)$ appropriately for a child-like vocal tract, and is dependent on age and sex. Since the magnitude of the lip opening can be similar for children and adults during speech production (e.g., Riely and Smith, 2003; Bunton *et al.*, 2013), the scaling function should reduce the cross-dimension more in the back portion of the vocal tract than in the front.

The approach used here is based on the hypothesis that regions along the VTL axis that are shortened relative to an adult baseline vocal tract will also be reduced in cross-section by a proportional amount. This can be accomplished by setting an initial scale factor to be the ratio of the absolute child VTL (at a target age) to the absolute adult tract length, VTL_β/VTL_α . The gradients of both the normalized adult and child length vectors generated by the length warping process (i.e., the length differences between successive elements in the length vectors) can then be used to weight the initial scale factor as a function of distance from the glottis. Thus, a scaling function κ can be defined as

$$\kappa(i) = \left(\frac{VTL_\beta}{VTL_\alpha} \right) \left(\frac{\nabla X_\beta(i)}{\nabla X_\alpha(i)} \right), \quad i = [1, N_x], \quad (4)$$

where $\nabla X_\alpha(i)$ and $\nabla X_\beta(i)$ are the gradients of the adult and child (at a target age) length vectors, respectively. It is noted that, based on the length warping process in Sec. IV A, the adult length vector is uniformly incremented from element to element, so its gradient in the denominator could be replaced with the increment value as a constant. The gradient notation is used, however, to maintain generality for possible nonuniform versions of the adult vector. A final operation is smoothing $\kappa(i)$ with a fourth order finite impulse response filter set to have a normalized cutoff frequency of 0.05. As an example, the scaling function shown in Fig. 6(b) was generated by applying Eq. (4) (and smoothing filter) to the length vectors shown previously in Fig. 6(a) (for a 4 yr old male) along with the corresponding values of VTL_α and VTL_β . It can be seen that $\kappa(i)$ gradually increases from a value of 0.54 at the glottis to 0.86 at the lips, thus reducing the cross-dimension values more in the back portion of the vocal tract than in the front.

C. Age-dependent cross-dimension scaling functions

Using the adult male vocal tract data as a baseline, the process described in Secs. IV A and IV B was used to generate length warping and scaling functions at one year increments from age 0 to 12 yrs, for both male and female, as well as for an adult female. The initial step was to determine the normalized β vectors from the growth curve set representing each selected age (see Figs. 4 and 5). These vectors are given in Table I, along with the absolute tract length values VTL_β . The entry for the adult male is blank because it is the α vector that represents the baseline adult male vocal tract (i.e., $\alpha = [0 \ 0.2639 \ 0.5988 \ 0.9180 \ 1]$ and $VTL_\alpha = 17.6$ cm).

The calculated scaling functions, for male and female, are shown in Figs. 7(a) and 7(b), respectively. The even years are indicated by thick lines and denoted with a label at the lip end, whereas the odd years are shown as dashed gray lines without labels. As noted previously for the 4 yr-old male, the scaling functions based on the male measurements [Fig. 7(a)] are fairly flat in the back half of the vocal tract, but increase with steeper slopes in the front half. The 0 yr scaling function has a more unique shape than the other ages, apparently due to the different distribution of segment lengths at the youngest age. Each yearly increment increases the magnitude of the scaling function (i.e., less reduction of the cross-dimension). The flat

TABLE I. Age and sex dependent β vectors and VTL values used to transform the adult male vocal tract model to sex- and age-dependent versions.

Age, χ (yrs)	β male						β female					
	A	B	C	D	E	VTL $_{\beta}$ (cm)	A	B	C	D	E	VTL $_{\beta}$ (cm)
0	0	0.2391	0.5581	0.8868	1	8.47	0	0.2457	0.5201	0.8969	1	8.43
1	0	0.2298	0.5320	0.8944	1	9.95	0	0.2257	0.5095	0.9015	1	9.56
2	0	0.2251	0.5285	0.8956	1	10.75	0	0.2163	0.5132	0.9043	1	10.36
3	0	0.2237	0.5308	0.8953	1	11.24	0	0.2123	0.5219	0.9066	1	10.93
4	0	0.2240	0.5341	0.8952	1	11.63	0	0.2111	0.5307	0.9079	1	11.38
5	0	0.2251	0.5375	0.8958	1	11.98	0	0.2115	0.5379	0.9086	1	11.75
6	0	0.2266	0.5410	0.8968	1	12.32	0	0.2129	0.5433	0.9089	1	12.07
7	0	0.2283	0.5445	0.8981	1	12.67	0	0.2152	0.5473	0.9091	1	12.37
8	0	0.2301	0.5480	0.8997	1	13.02	0	0.2181	0.5501	0.9092	1	12.66
9	0	0.2321	0.5516	0.9012	1	13.38	0	0.2217	0.5523	0.9094	1	12.95
10	0	0.2341	0.5551	0.9027	1	13.75	0	0.2257	0.5541	0.9096	1	13.25
11	0	0.2363	0.5587	0.9041	1	14.13	0	0.2298	0.5555	0.9098	1	13.55
12	0	0.2386	0.5624	0.9055	1	14.51	0	0.2337	0.5569	0.9101	1	13.84
Adult							0	0.2412	0.5666	0.9090	1	15.55

scaling function with a constant value of 1.0 for the adult male simply reflects the fact that it was the baseline from which all other ages, male and female, were derived. The female scaling functions in Fig. 7(b) for ages 2 yrs to adult all slope upward in the back half of the tract beginning with values that are slightly smaller than those of the male. For the 0 and 1 yr cases, scaling functions slope downward in the back portion of the tract. All the functions then rise rapidly to about the middle of the oral cavity where the scaling values are comparable to those calculated for the male data, and then either become flat or drop somewhat at the lips. The adult female scaling function has a shape that is similar to the others, but the overall values are greater than for all other ages. It is noted that if adult *female* vocal tract measurements were used as the baseline data, the scaling function for the adult female would be a flat line at 1.0, and all other scaling functions, including the adult male, would be shaped according to those data.

D. Transformation of the vocal tract model components

The adult vocal tract area function model described in Sec. II can now be transformed to a child-like version. To demonstrate, the neutral diameter function $\Omega(i)$ (see Fig. 1 and Table III in the Appendix) is plotted in Fig. 8(a) relative to the normalized length axis X_χ [Eq. (3)]. The thin gray line shows the same $\Omega(i)$ function, but plotted against the warped axis X_β for the 4 yr-old male from Fig. 6(a), and shows a shift of the overall shape toward the glottis. The effect of the length warping combined with application of the scaling function from Fig. 6(b) is shown by the dashed-dotted line [$\kappa(i)\Omega(i)$ vs $X_\beta(i)$], and results in a large reduction of diameter values along the length of the tract, more so in the back half than in the front. The transformation is completed by resampling the warped and scaled version of the $\Omega(i)$ function so that the x axis contains equidistant intervals. This was carried out with the same piecewise cubic interpolation approach used in Sec. IV A. Finally, an absolute length vector can be generated as

$$x(i) = \sum_{z=1}^i L_\chi(z), \quad i = [1, N_x], \quad (5)$$

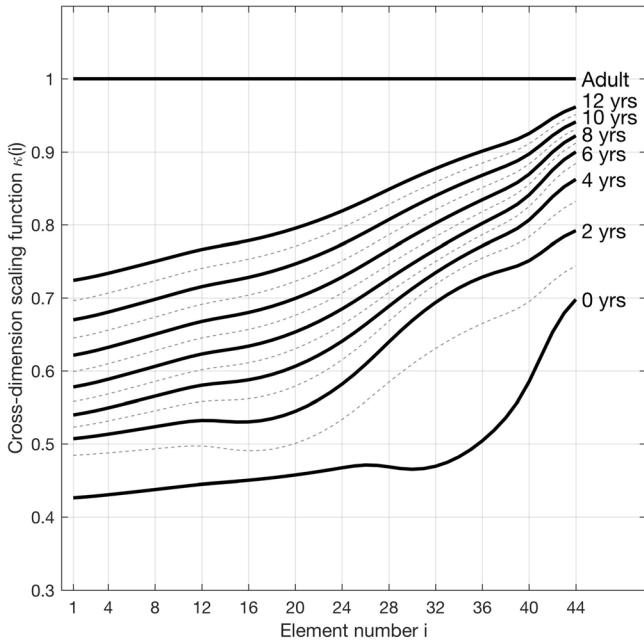
where $L_\chi(i) = \text{VTL}_\beta/N_x$, and represents length of each vocal tract element for a given age ($N_x = 44$ in this study). The adult male $\Omega(x)$ and the transformed 4 yr-old male version are shown in Fig. 8(b) projected onto their respective absolute length axes.

Following this same process, the $\Omega(i)$, $\phi_1(i)$, and $\phi_2(i)$ components of the adult model were modified across age and sex to generate a collection of child-like vocal tract models. The results are shown in Fig. 9 where, in all plots, the thick black line represents the original adult male model, the thick dashed-dotted line is the transformation to the youngest age (0 yr), and all other lines represent the progression through ages 4, 8, and 12 yrs (other years were not shown to maintain clarity in the plots). The additional thick gray line in the right column plots represents the transformation to an adult female model. The origin of the length axis was offset to be the junction of the HPSP (landmark “B”) so that the growth in both the front and back portions of the vocal tract could more easily be seen. Contraction in both length and cross-dimension is apparent for all model components, and there are some observable differences between the male and female versions. For example, at the youngest age the neutral function, $\Omega(x)$, is slightly larger in the back half for the male than the female, but opposite in the front half. The model components, $\Omega(i)$, $\phi_1(i)$, and $\phi_2(i)$, derived for each of the ages shown in Fig. 9 are provided in numerical form in Tables III and IV in the Appendix.

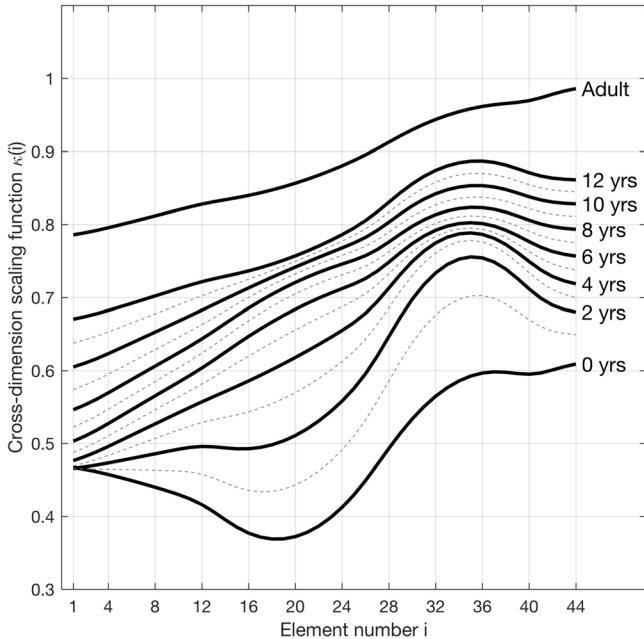
Using the model components derived for each age represented in Fig. 9, the vowel /u/, as defined by the $[q_1, q_2] = [1.68, -1.87]$ coefficients given in Table II (see the Appendix), was generated with

$$D(i) = \Omega(i) + q_1\phi_1(i) + q_2\phi_2(i), \quad (6)$$

which is a modified version of Eq. (1) that produces a diameter function rather than cross-sectional areas along the tract length. This was done so that the vocal tract shape could be shown in a pseudo-midsagittal view (cf. Story *et al.*, 2001; Story, 2013) in which diameters from glottis to lips are plotted



(a) Male $\kappa(i)$

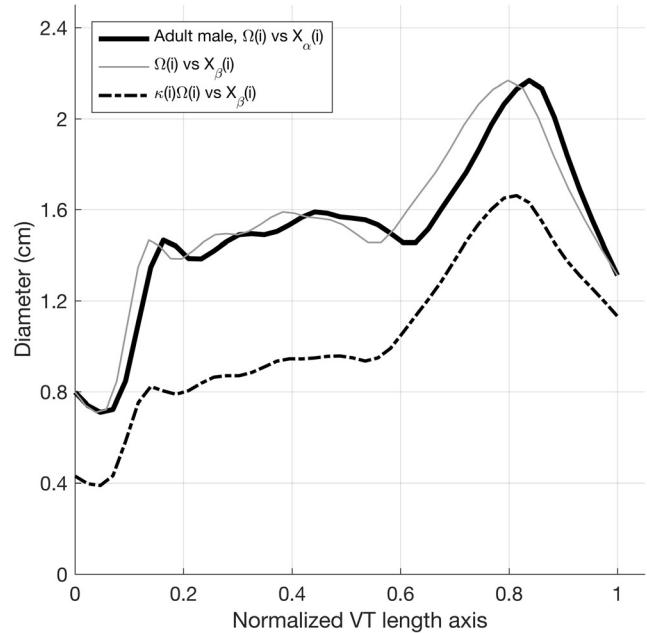


(b) Female $\kappa(i)$

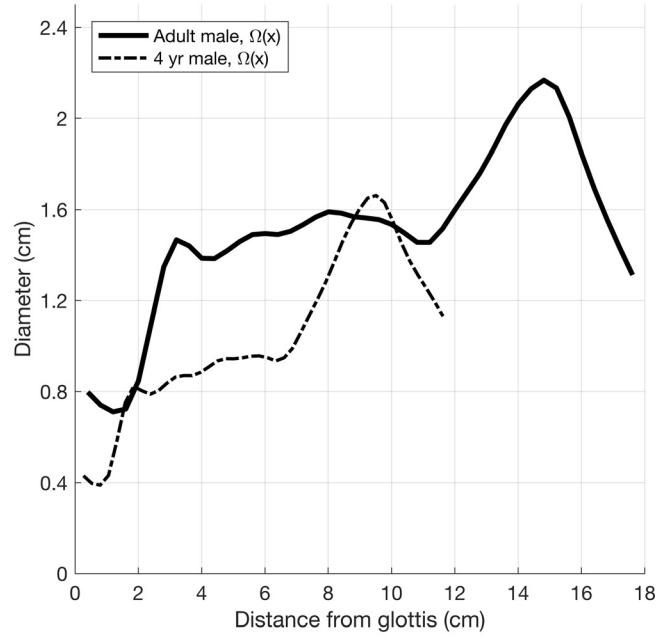
FIG. 7. Cross-dimension scaling functions for ages 0–12 yrs and adult plotted as a function of element number, and ordered from glottis to lips: (a) male and (b) female.

perpendicular to a vocal tract centerline profile.² The /ʊ/ vowel was chosen as a demonstration case because it requires contributions of each mode that are of similar magnitude.

The resulting age and sex dependent tract configurations are shown in Fig. 10 where the origin is again located at the HPSP junction, age increases from bottom to top, and male and female are in the left and right columns, respectively. The β vector points for each age are indicated by the open circles (excluding the glottal and lip points), and the division into four vocal tract regions are shown in the same manner as in Fig. 5.



(a)



(b)

FIG. 8. Demonstration of applying calculated length warping and cross-dimension scaling functions for a 4 yr-old male to the adult male $\Omega(i)$ from Fig. 1(a). (a) $\Omega(i)$ vs X_z , $\Omega(i)$ vs X_β , and $\kappa(i)\Omega(i)$ vs X_β are shown with a normalized VTL axis. (b) $\Omega(x)$ for the adult male and 4 yr-old plotted with an absolute length axis.

The overall VTL for each age is indicated in the upper right corner of each panel. Although this representation is not an accurate depiction of the true midsagittal plane, it does allow for projection of an otherwise abstract area function onto a more anatomically-relevant perspective indicating how expansions and constrictions of the airspace are distributed along the vocal tract axis. This perspective shows the age-dependent increase in length and vertical descent of the pharyngeal portion of the vocal tract, a lesser increase of the oral cavity

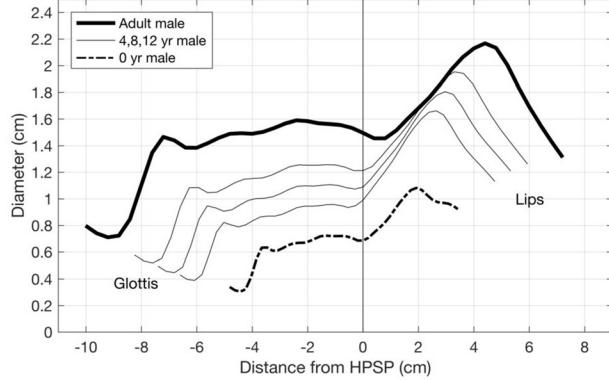
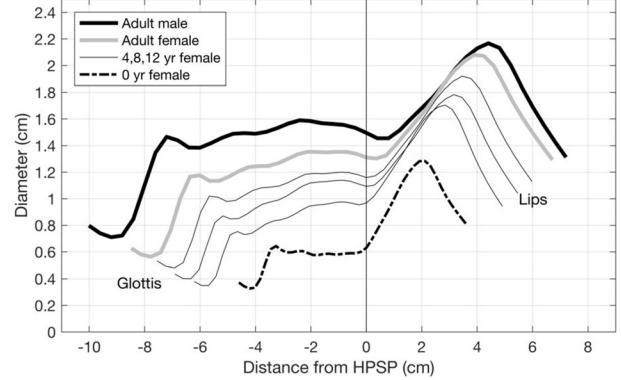
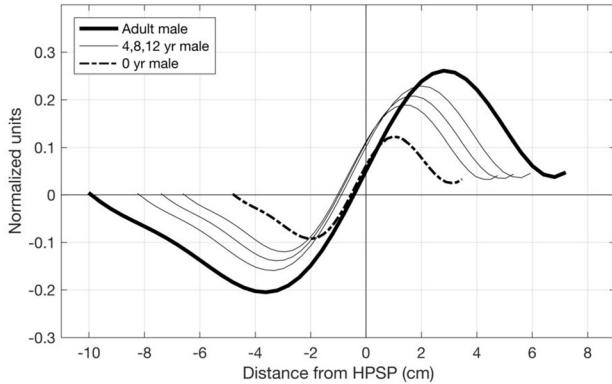
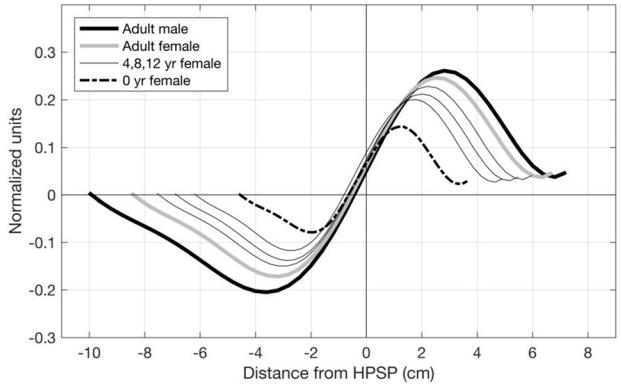
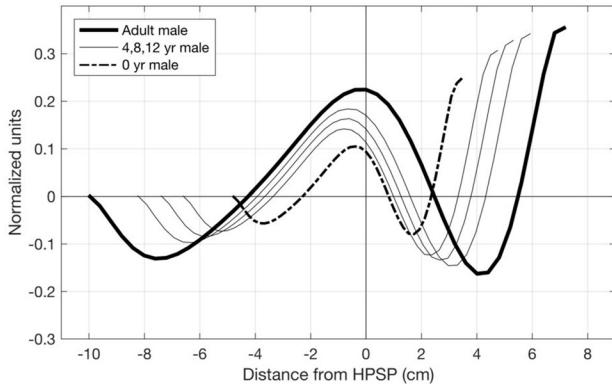
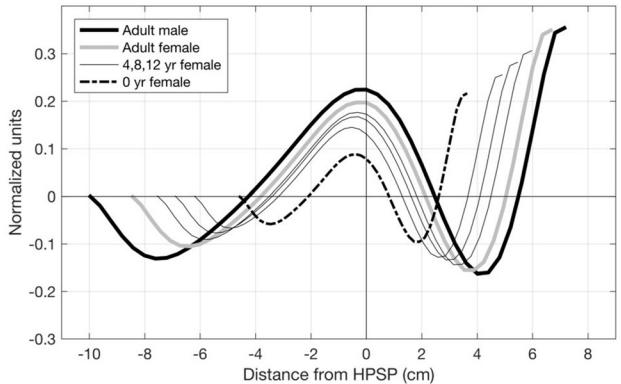
(a) $\Omega(x)$, male(b) $\Omega(x)$, female(c) $\phi_1(x)$, male(d) $\phi_1(x)$, female(e) $\phi_2(x)$, male(f) $\phi_2(x)$, female

FIG. 9. Length warped and scaled versions of the components of the vocal tract model shown for ages 0, 4, 8, 12 yrs and adult where the left column represents male and the right column female. The origin of each plot is the HPSP junction (landmark B). (a) and (b) $\Omega(x)$; (c) and (d) $\phi_1(x)$; (e) and (f) $\phi_2(x)$.

length, and the nonuniform increase in cross-dimension magnitude. For the 0 yr case, the cross-dimension of the male oral cavity region is smaller than for the female, whereas the male and female configurations are quite similar for the 4 and 8 yr cases, after which the male tract becomes longer in the pharyngeal region than the female.

V. PLAUSIBILITY OF TRANSFORMED ADULT-TO-CHILD VOCAL TRACT SHAPES

The plausibility of vocal tract configurations produced by the transformation process was assessed in two ways. First,

pseudo-midsagittal plots, similar to those in Fig. 10, were projected onto actual midsagittal images of two children. This provides a visual, qualitative assessment regarding how well the hypothetical tract shapes appear to “fit” an age-matched vocal tract image. The second assessment was based on comparing hypothetical area functions to cross-sectional area measurements obtained for children and adults using acoustic pharyngometry. Although this technique is an indirect means of measuring area values, it provides at least some quantitative validation of the transformation process at the level of the area function.

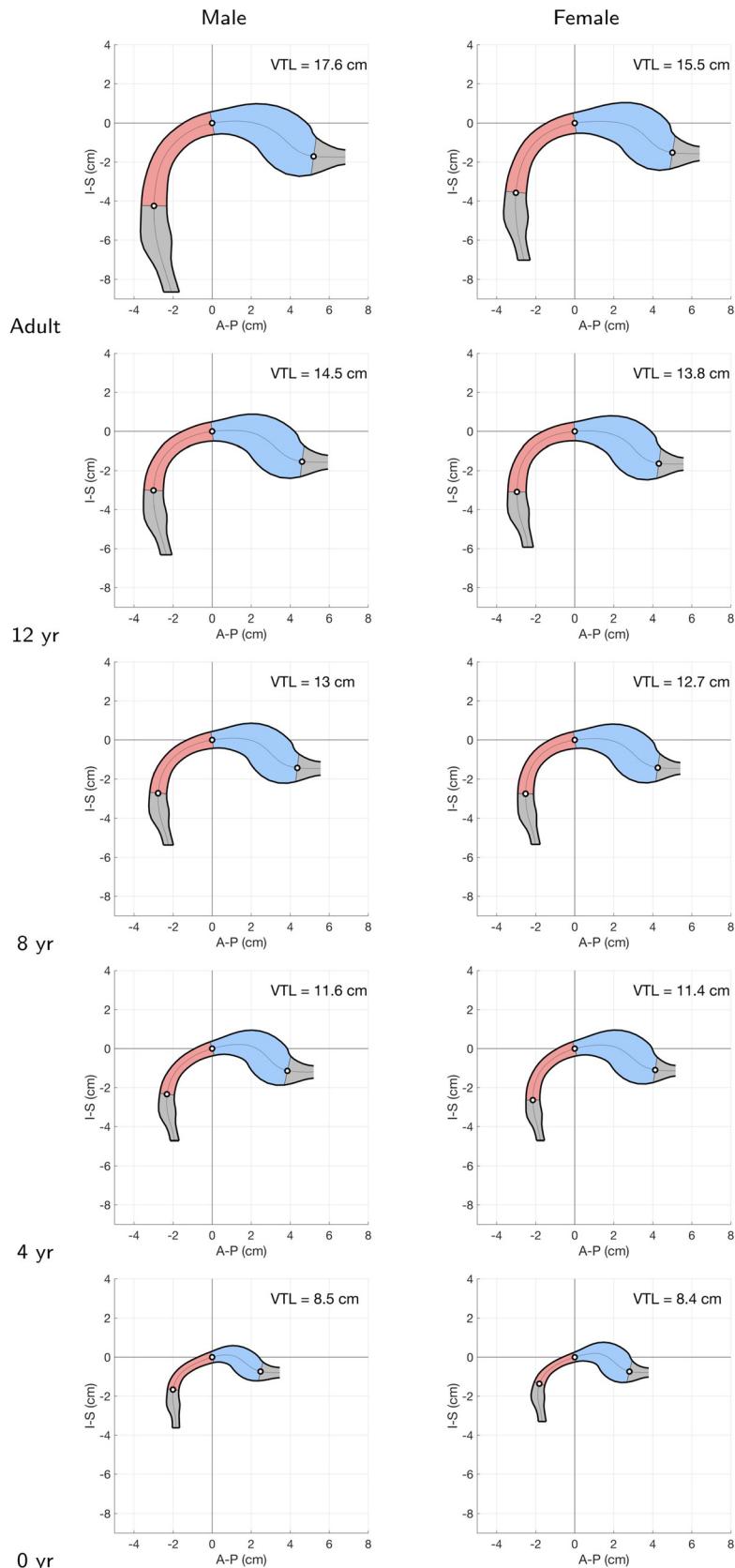


FIG. 10. (Color online) Pseudo-midsagittal plots of the hypothetical /o/ area functions generated by the length warping and scaling process for ages 0, 4, 8, 12 yr and adult based on the male and female segment polynomials (Figs. 4 and 5). Age increases from bottom to top. In each panel, the HPSP junction (landmark B) is the origin of the axes and the open end on the right is the anterior aspect of the lips. The open circles indicate the points B, C, D, for each age. The colors of each segment are the same as those in Fig. 5.

A. Projection of hypothetical vocal tract shapes onto midsagittal images

From the image sets used to extract the anatomic measurements, midsagittal sections were selected for one male (3 yrs, 11

months) and one female (1 yr, 9 months) participant. The specific images were chosen such that the mouth was open and that they represent fairly young ages. As an approximate match, pseudo-midsagittal plots like those in Fig. 10 were generated based on the transformation process for a hypothetical 4 yr-old

male and a 2 yr-old female producing a neutral vowel and /i, a, u/. The pseudo-midsagittal plots were resized to match the pixel calibration of the midsagittal images and projected onto them as shown in Fig. 11. Each vocal tract shape was rotated to align with the appropriate anatomical landmarks (points A, B, C, and D), and made to be slightly transparent so that the features present in the images can be seen through each hypothetical vocal tract shape. Because the images were collected during quiet breathing, the velum is in a lowered position for both the male and female. Nonetheless, the projected vowel shapes include regions of constriction and expansion that would appear to be possible productions by both the male and female in the images. The /i/ vowel is perhaps closest to the shape produced during quiet breathing where both the pharyngeal and oral cavity are fairly well matched to the images.

The midsagittal image for the male in the quiet breathing condition was part of a larger volumetric image set which allowed for measurement of the cross-sectional area within the oral and pharyngeal cavities. The black arrows on each of the images in the upper row of Fig. 11 point to a location in the pharynx where the area was measured to be 2.0 cm^2 , and an oral cavity location with a measured area of 0.7 cm^2 . In comparison, the area at the same pharyngeal location in the model-generated vowel shapes is $0.6, 1.5, 0.15$, and 1.3 cm^2 for the neutral, /i/, /a/, and /u/ vowels, respectively. The 1.5 cm^2 area at this location for the projected /i/ tract shape is closest to the 2.0 cm^2 area measured from the image set, and accordingly its equivalent diameter is somewhat smaller than the cross-distance of the midsagittal image. The areas at the oral cavity location in the model-generated vowel shapes are $1.4, 0.1, 3.3$,

and 0.7 cm^2 for the same four vowels. Here, the area of 0.6 cm^2 in the projected /u/ vowel is just slightly less than the area measured from the image set, and corresponds to near equality of the diameter and midsagittal cross-distance measures. The cross-sectional area values produced by the model are clearly in the same range of the measured area values from the image set, suggesting that the vocal tract area functions generated by the child-like versions of the model are plausible configurations of these vowels.

B. Comparison of area functions measured with acoustic pharyngometry to those based on the vocal tract growth model

Transformation of the vocal tract model was also evaluated by comparison to area measurements based on acoustic pharyngometry, a technique in which acoustic reflections emitted into the vocal tract via a mouthpiece are used to estimate cross-sectional areas along the VTL. For a detailed description, see Vorperian *et al.* (2015); protocols on data collection and analysis are available at <http://www.waisman.wisc.edu/vocal/resources>. Although the acoustic pharyngometer generates only indirect measures of cross-sectional areas in the oral and pharyngeal portions of the vocal tract during slow exhalation, not production of target vowels, it does provide a means of comparing adult vocal tract areas to those of children of nearly any age. Vocal tract cross-sectional area reconstructions were obtained with the Eccovision acoustic pharyngometer (Sleep Group Solutions, Hollywood, FL) for an adult male and eight children (4 male, 4 female) ranging in age from 4 to 7 yrs.

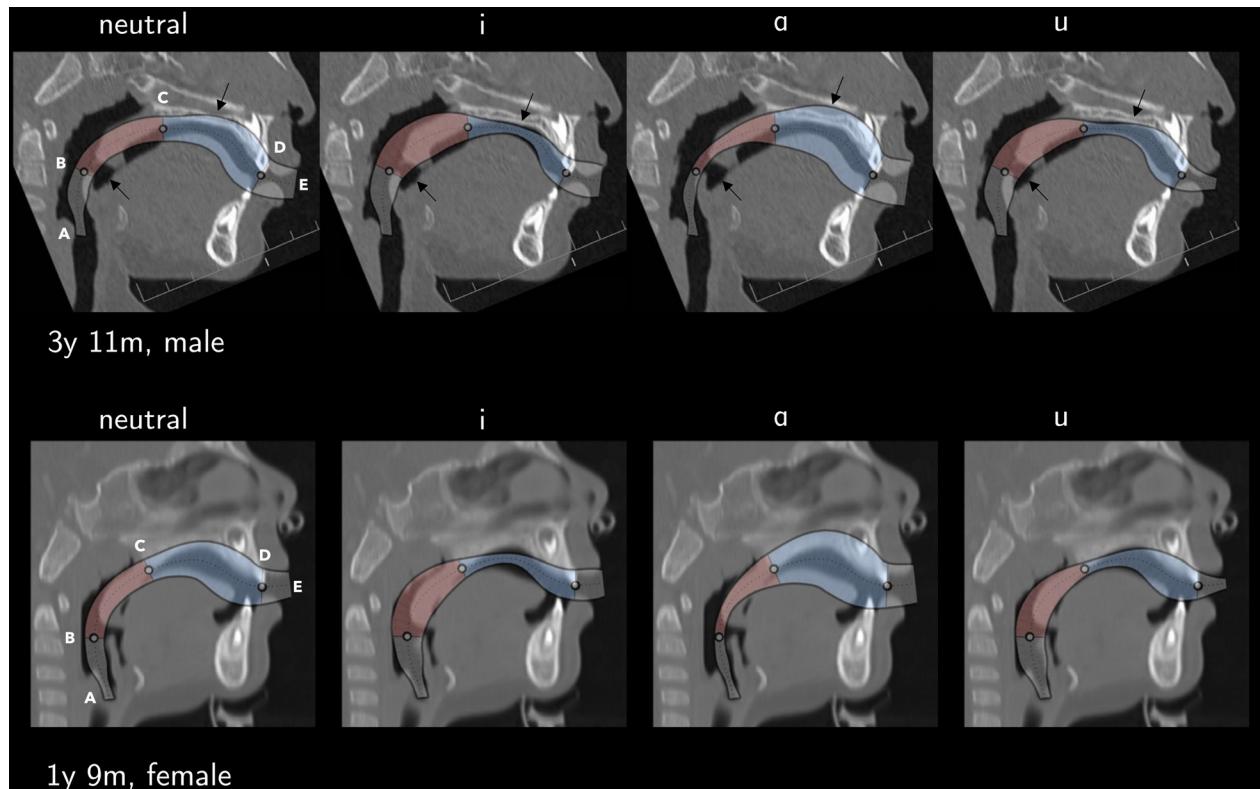


FIG. 11. (Color online) Pseudo-midsagittal plots of hypothetical area functions for a neutral vowel and /i, a, u/ projected onto CT images of 3 yr 11 month old male (top row), and a 1 yr 9 month old female (bottom row). The black arrows shown on the images in the upper row point to pharyngeal and oral locations at which the cross-sectional area was measured (see text for explanation).

Vocal tract area reconstructions obtained with the Eccovision acoustic pharyngometer are shown in Fig. 12, and are denoted in the legends as “APh.” In both panels, the same adult male vocal tract shape is plotted as the thick black line, whereas the gray dashed lines are the 4, 5, 6, and 7 yr old males in Fig. 12(a) and females of the same ages in Fig. 12(b). The mouthpiece of the Eccovision system dictates that the cross-sectional area at the termination of all vocal tract shapes (the lingual aspect of the incisors) is approximately 3 cm^2 , as can be seen in both Figs. 12(a) and 12(b). Thus, for this figure it is convenient to plot the area functions with respect to the distance from the mouthpiece rather than referenced to the glottis or HPSP. The points labeled “OPJ” are the approximate location of the oropharyngeal junction. Additionally, because there were no systematic differences across age within the male and female groups, age is not differentiated by line type in the plots, but the OPJ symbols are unique to each age. There are, however, large adult versus child differences in area along the entire tract length, except near the lips as expected due to the influence of the mouthpiece.

To test the growth transformation process, the adult area function obtained with the Eccovision acoustic pharyngometer was modified by applying the scaling function, $\kappa(i)$, appropriate for a male and female at an age of 6 yrs (see Fig. 7), and then compared to the *actual* APh area functions obtained for children (ages 4 to 7). Because the growth transformation process was designed to be applied to the vocal tract model in the equivalent diameter domain [i.e., prior to the squaring operation in Eq. (1)], the APh area function was converted to a diameter function prior to multiplication by $\kappa(i)$ and then returned back to the area domain. In addition, the four anatomic landmarks were not available in the APh data, so the length warping could not be applied directly to the adult APh area function. Instead, the length axis computed for 6 yr olds, male and female, based on the anatomic data, was used for plotting the transformed APh area functions.

The solid gray line in each panel of Fig. 12 is the transformation of the adult male APh function to a hypothetical version of a 6 yr-old child, male in the upper panel and female in the lower. This age was chosen simply as a demonstration; hypothetical area functions could be generated at any of the other ages within the range of the APh data and would be slightly different in the cross-sectional area and VTL. The hypothetical, model-generated male area function in Fig. 12(a) is quite similar in overall shape to the other APh area functions for the 4–7 yr-old males except at the lips (i.e., the growth transformation does not account for Eccovision mouthpiece). Along most of the tract length, the cross-sectional area passes within the range of the APh child data. This is easily observed within the ellipse containing the OPJ points located at about -7 cm from the lips. In Fig. 12(b), the model-generated female area function is, for the most part, similar to the APh area functions for 4–7 yr females, but is well below one APh area function with area values that exceed 2.5 cm^2 at a location of about -8 cm from the lips. This belongs to the 5 yr old female and, relative to the APh data for the other children, seems to be an outlier. Otherwise the hypothetical areas are mostly within the range of the APh female data along the entire vocal tract, except at the lips. These plots indicate that the growth

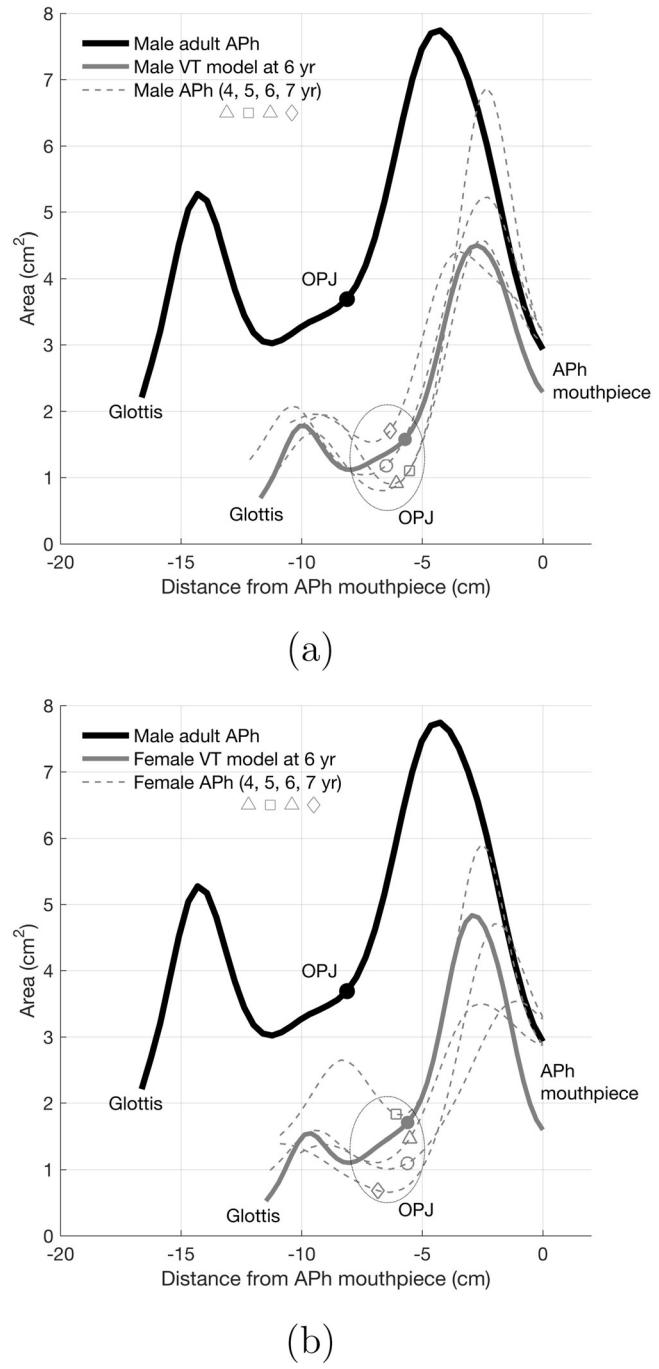


FIG. 12. Area functions obtained for an adult male and eight children using the Eccovision acoustic pharyngometer, and projections of the adult male area function to a 6 yr-old male and female based on the cross-dimension scaling functions from Fig. 7. Each area function is plotted with respect to the distance from the pharyngometer mouthpiece which is located at the lip end of the vocal tract, hence the x axis is negative. The points labeled OPJ denote the approximate location of the oropharyngeal junction. In the legends, APh indicates acoustic pharyngometry data. (a) The thick black line is the adult male vocal tract shape based on APh cross-sectional areas obtained during slow exhalation. The gray dashed lines are APh-based vocal tract area functions measured from males 4–7 yrs old where the symbols denoting the OPJ are different for each age (note that the symbols are shown in the legend below the corresponding age). The thick gray line is a projection of the adult APh area function to a 6 yr-old male based on applying the appropriate scaling function from Fig. 7(a). (b) The thick black line is the same adult male vocal tract shape as shown in (a). The gray dashed lines are APh-based vocal tract area functions measured from females 4–7 yrs old, where again the symbols denoting the OPJ are different for each age. The thick gray line is a projection of the adult area function to a 6 yr-old female based on applying the appropriate scaling function from Fig. 7(b).

transformation process is capable of projecting an adult vocal tract area function to one that is appropriate for a given age.

VI. AGE-DEPENDENT VOWEL SPACE BASED ON TRANSFORMATION OF THE VOCAL TRACT MODEL

Using the length-warped and scaled versions of the vocal tract model (Fig. 9 and Tables III and IV in the Appendix), a vowel space plot, like the one shown in Fig. 1(c), was generated at each of the ages 0, 4, 8, and 12 yrs, as well as for the adult stage, for both male and female. The coefficient grid used to generate each vowel space was identical to Fig. 1(b) whose boundaries are defined by the q_1 and q_2 values in Table II (see the Appendix), i.e., exactly the same coefficients were used regardless of age and sex. The ten vowel space plots are shown in Fig. 13 and are arranged similarly to Fig. 10 where age increases from bottom to top, and male and female are in the left and right columns, respectively. Each gray mesh shown in Fig. 13 is comprised of the resonance frequencies, f_{R1} and f_{R2} , calculated for 1600 area functions constructed with the vocal tract model for a given age and sex. The irregular termination of the mesh in the upper left corner of most cases is due to removing a few points in which the second and third resonances were in such close proximity that the third resonance was misidentified as the second. As a reference, the four solid dots plotted on top of each mesh are, in clockwise order beginning in the upper left, the resonance frequencies corresponding to the coefficients for the /i, a, æ, u/ vowels in Table II (see the Appendix). Even though the size and shape of the mesh changes at each age increment, its one-to-one relation to the coefficient grid [see Fig. 1(b)] is maintained at all ages.

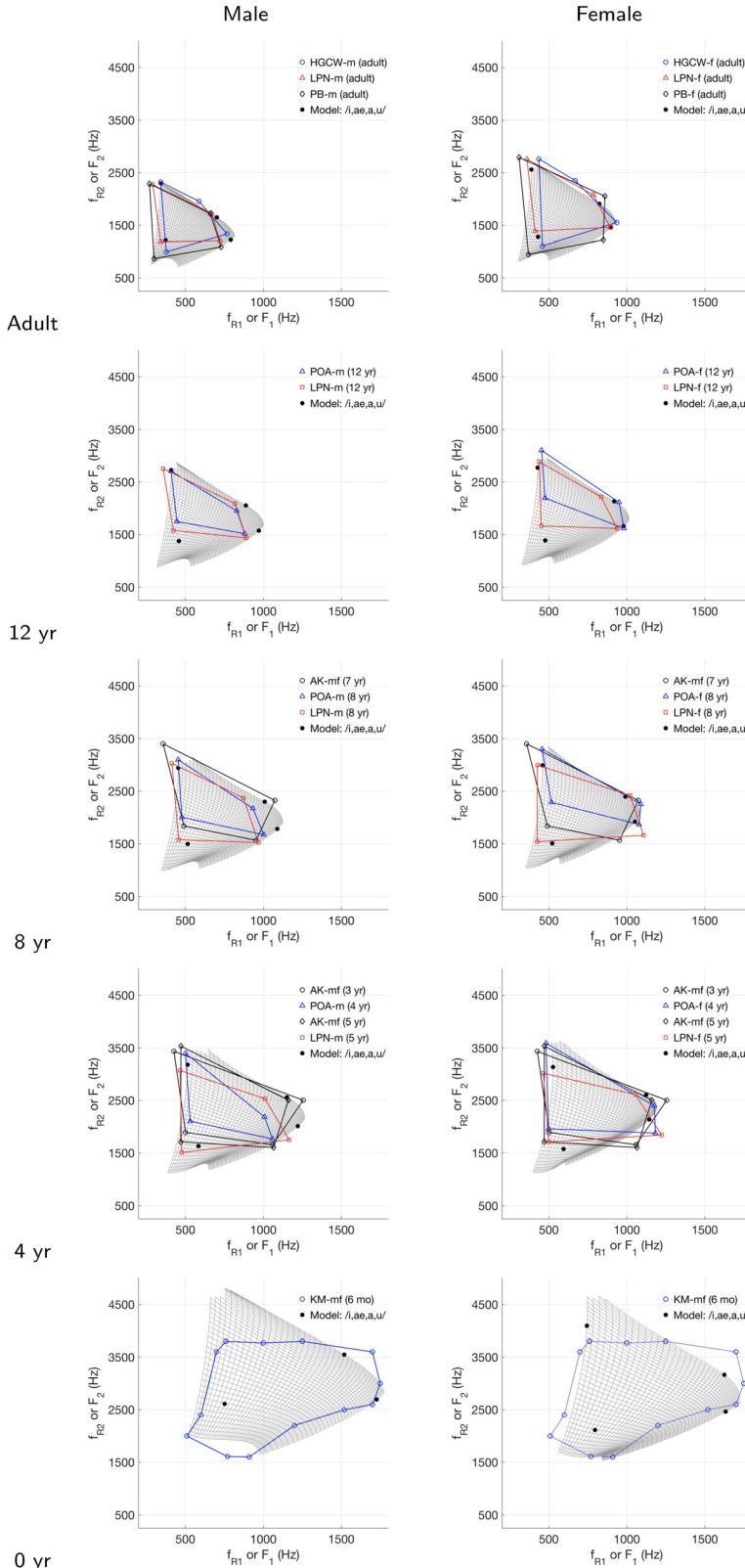
A. Comparison of simulated vowel space plots to published formant measurements

To assess the ability of the age-dependent vocal tract model to produce realistic resonance frequencies, measured F_1 and F_2 formant values obtained from several reported studies were used for comparison to the model-generated meshes in Fig. 13. None of the studies contained a data set that spans the entire age range represented in Fig. 13, thus a combination of data sets across the studies was used for each age increment. With the exception of the 0 yr case, data for only the vowels /i, æ, a, u/ were used for comparison. Formants measured from a variety of vowel-like sounds produced by six-month old children (both male and female) were estimated from a figure in Kent and Murray (1982, p. 357); they will be referred to as "KM-mf" in the vowel space plots. Assmann and Katz (2000) reported formant measurements of 30 children of ages 3, 5, and 7 yrs; the distribution of males and females was not reported, hence the same formant values will be used for both the male and female cases and will be denoted as "AK-mf." Perry *et al.* (2001) measured the formant frequencies of 20 children (10 M; 10 F) in groups comprised of 4, 8, 12, and 16 yr olds. The data of the 4, 8, and 12 yr-old children, male and female, were used for comparison to the model generated vowel spaces corresponding to the same ages, and will be denoted as "POA-m" and "POA-f." Based on analysis of 436 children (229 M; 207 F) and 56 adults (29 M; 27 F), Lee *et al.* (1999) (hereafter, "LPN") reported

formant frequencies in yearly increments ranging from 5–18 yrs as well as the adult stage. The data for the 5, 8, 12 yr-olds and adults were compared to the model-generated cases of similar age. These data were reported separately for male and female talkers and are indicated by "LPN-m" or "LPN-f" in the legend of the relevant plots. For an additional comparison to the model-generated adult vowel spaces, formant data for males and females were used from Peterson and Barney (1952) (denoted as "PB-m" and "PB-f") and Hillenbrand *et al.* (1995) (denoted as "HGCW-m" and "HGCW-f").

In the bottom row of Fig. 13 are the male and female vowel space plots generated for the 0 yr (0–6 months) stage. For the male, the frequencies of the first and second resonances range from 520 to 1780 Hz and from 1660 to 4840 Hz, respectively. For the female, the first resonance has a minimum of 564 Hz and a maximum 1725 Hz, and the second resonance ranges from 1530 to 4684 Hz. Formant data reported for 6 month-old children (both male and female) in Kent and Murray (1982) has been plotted on top of both the model-generated meshes. With the exception of a few points in the upper right and lower left portions of the vowel spaces, the formant data are contained within both the male and female meshes, suggesting that the model generates reasonable vocal tract configurations for this age. The point outside the mesh in the upper right portion of each vowel space could likely be generated if the mouth opening of the model were allowed to become larger and the VTL shortened as might be the case for production of an /æ/-like vowel by a 6 month old. It can be noted that for this case, the formant data do not represent target vowels, but are simply based on a variety of sounds produced by the 6 month-old children in the Kent and Murray (1982) study. The vowel space mesh generated by the model can be viewed in the same way. It represents possibilities for sound production by the vocal tract at this early age.

The 4 yr-old male and female cases are shown in the fourth row of Fig. 13 and indicate that calculated resonance frequencies are shifted downward in frequency along both axes, and the mesh is reduced in overall size, as expected due to increased tract length. The frequency ranges of the first and second resonances for the male are [390, 1270] and [1120, 3490] Hz, respectively, whereas they are [396, 1120] and [1148, 3647] Hz for the female. The AK data of the 3 and 5 yr-old children, the POA data for 4 yr-olds, and the LPN data for 5 yr-olds are plotted on top of each resonance frequency mesh for the 4 yr-old cases. Even with the wide variation of the measured formant data, the polygons formed by the four vowels in each data set are all nearly contained within the vowel spaces generated by the 4 yr-old versions (male and female) of the vocal tract model; the /i/ and /a/ formants are just outside the meshes. The formant data for the AK 7 yr-olds, and 8 yr-olds from the POA and LPN studies are all similarly contained by the model-generated meshes for the 8 yr-old stage (third row of Fig. 13), but again some of the vowel data lies just outside the meshes. The 12 yr-old cases for both male and female (fourth row) indicate that the model-generated meshes have been contracted in overall size, but are not yet adult-like. The LPN formant data for male and female 12 yr-olds fits well within the model-generated mesh, as does the POA data for 12 yr-old



males. The exception is the POA data for 12 yr-old females where the /i/ and /æ/ vowels extend just above the mesh.

The adult vowel space plots are shown in the top row of Fig. 13. The male version is the same as shown previously in Fig. 1(c), but is now plotted along with /i, æ, a, u/ formant data from the studies HGCW, LPN, and PB. Other than two of the /i/ vowels, the model-generated mesh contains the formant data; in fact, the ranges of both the mesh and the measured

FIG. 13. (Color online) Vowel space plots (gray meshes) generated by the age-dependent vocal tract model for a male (left column) and female (right column). The plots are arranged identically to Fig. 10 where age increases from bottom to top. As a reference, the four solid (black) dots plotted on top of each mesh are, in clockwise order beginning in the upper left, the resonance frequencies corresponding to the coefficients for the /i, a, æ, u/ vowels in Table II (see the Appendix). The data points plotted as polygons are from several studies that have reported measured formant frequencies across a range of ages. The legends contain abbreviations for the cited studies (see text for details). For all polygons, except those in the bottom row, the four vertices represent the vowels /i, a, æ, u/ in clockwise order, beginning in the upper left corner. Because the gray mesh in each subplot is based on calculated vocal tract resonance frequencies, and the discrete data points are measured formant frequencies, the horizontal and vertical axes are labeled as f_{R1} or F_1 and f_{R2} or F_2 , respectively (footnote 1).

formants are nearly the same. The adult female case is similar, where a few of the measured vowels are just outside the mesh.

In summary, the model-generated vowel space meshes for both male and female contract in overall size and shift downward along both the horizontal and vertical axes as age increases from bottom to top in Fig. 13. Although this is expected due to the increase in VTL that occurs during development, the non-uniform length changes of the four regions of the vocal

tract, along with corresponding cross-dimension scaling also generate changes in the shape of the mesh from one age increment to the next, as well as male relative to female. For example, the lower left portion of the mesh in the 0 yr female case is pulled inward to a greater degree than the male, and results in a region extending roughly from [800, 1600] to [1100, 2000] Hz with a higher density of resonance pairs. Similar high density regions can be seen in both the male and female meshes at the other ages, but not the adult stage. This effect is presumably due to the interaction of the length-warping and scaling process with the idiosyncrasies of the particular vocal tract on which the model is based.

B. Simulated vowels based on the length-warped and scaled vocal tract model

A simulated vowel was produced based on each of the area functions corresponding to the solid black dots in each model-generated mesh, male and female, of Fig. 13. These represent the coefficients of the /i, æ, ɑ, u/ vowels in Table II (an example of these vocal tract shapes for a 4 yr-old male were shown in the top row of Fig. 11). The simulations were produced with the TubeTalker airway modulation model (Story, 2013), a speech simulator whose voice source and vocal tract parameters can be scaled to be appropriate for various ages. Although the details of TubeTalker have been reported elsewhere and are beyond the scope of this article, it is noted that the voice source is a kinematic model of the medial surfaces of the vocal folds whose frequency of vibration can be precisely specified. For these synthetic vowels, the fundamental frequency (f_o) for each vowel was set to rise and fall around a nominal value that was age dependent based on data reported by Eguchi and Hirsh (1969, Table II, p. 24). The f_o values are the same for males and females up to age 8 yrs, and then became sex-specific. The duration of each vowel is 0.5 s. Each sample has been low-pass filtered with a cutoff frequency of 6000 Hz to preserve only the portion of the spectrum that can be represented by one-dimensional wave propagation.

The vowels are presented for listening in two audio files, Mm. 1 and Mm. 2, where the first represents the male vocal tract models and the second file represents the female vocal tract models. Each audio file arranged such that the /i, æ, ɑ, u/ vowels at a given age are played in successive blocks separated by a narrator's voice announcing each age. These audio samples are made available simply to provide an informal experiential means of demonstrating the output of the age-dependent vocal tract model.

Mm. 1 Audio file. This is a file of type mp3 (400 Kb).

Mm. 2 Audio file. This is a file of type mp3 (400 Kb).

VII. DISCUSSION

The length warping and cross-dimension scaling process developed in this study is essentially a reverse-growth model of the vocal tract. That is, a model that projects an adult vocal tract "backward" in time to a younger age while taking into account prepubertal sex-differences in vocal tract

morphology. The results indicate that the model produces reasonable child-like vocal tract shapes and vowel spaces, providing support for using it as a tool to explore characteristics of speech development.

The plausibility of vocal tract configurations generated for various vowels was demonstrated by direct comparison to mid-sagittal images obtained of children in a quiet breathing condition. In a qualitative sense, the expansions and constrictions that characterize the set of model-generated vowels can be seen to "fit" into the pharyngeal and oral cavity spaces exhibited in images of real children. Cross-sectional area measurements based on one set of images were also shown to be well aligned with areas produced by the model. This was by no means an exhaustive test, but does demonstrate that the warping and scaling process transforms an adult vocal tract model to one that can be considered "child-like" with regard to anatomic constraints. Specifically, the dependence of the cross-dimension scaling on the warping of the VTL axis non-uniformly attenuates the epilaryngeal and pharyngeal regions of the vocal tract model relative to the oral cavity and lip regions. This limits the magnitude of possible expansions and constrictions in the back half of the vocal tract to a greater degree than in the front half, effectively attenuating some of the adult-like degrees of freedom for shaping the tract. Additional support was provided by vocal tract area measurements obtained with an acoustic pharyngometer (APh). It was shown that an adult male APh area function could be transformed, by application of the male and female cross-dimension scaling functions, to child-like versions (male and female) comprised of area values similar to those of the APh area functions obtained from 4–7 yr-old children.

Plausibility of the vocal tract model was also established by comparison of the model-generated vowel space plots (f_{R2} vs f_{R1}) across the ages of 0–12 yrs and adult, to formant data reported in the literature. With the exception of the youngest age, the $[f_{R1}, f_{R2}]$ ranges of the modeled vowel spaces were large enough to contain most of the $[F_1, F_2]$ data points at each age for both male and female. The vowel spaces could likely be expanded to encompass all of the formant data by implementing variable VTL to account for actions such as lip protraction and retraction rather than maintaining a constant value for area function reconstructions (Mokhtari *et al.*, 2007; Story, 2009). It is also important to note that the child-like and adult female versions of the vocal tract model were derived from a particular adult male vocal tract (i.e., Story *et al.*, 1996). Any speaker-specific aspects of the original vocal tract model will have an effect on the characteristics of the vowel space meshes at each age, but if the model were based on a set of area functions obtained from a different talker (e.g., Story, 2005b), somewhat different vocal tract configurations and vowel spaces would likely be generated. Furthermore, the cognitive and motor control influences of the original adult talker are, to some degree, also embedded in the model and projected backward in time. Thus, the derived area function models generated by the transformation process do indeed produce *possible* vocal tract configurations, but they are dependent on a level of cognitive and motor development that may not be fully developed at a given age. In addition, acoustic side branches such as the piriform sinuses and trachea were not present in the calculations of resonance frequencies, but if included would

potentially have the effect of shifting the main vocal tract resonances either up or down in frequency depending on the degree of coupling, location, and size of the side branch structure, and configuration of the vocal tract itself (cf. Dang and Honda, 1997; Stevens, 2000; Pruthi *et al.*, 2007; Honda *et al.*, 2010; Delvaux and Howard, 2014; Vampola *et al.*, 2015). Investigation of the effect of side branch coupling on resonance frequencies and bandwidths of small vocal tracts is one of the next steps needed to better understand the acoustic characteristics of children's speech production.

The acoustic assessment of the vocal tract models in this study was limited to only the first two resonances, f_{R1} and f_{R2} , to maintain focus on the one-to-one mapping that exists between them and the components of the vocal tract model. Higher frequency resonances exist, however, for any given area function and are readily available from frequency response calculations. The simulated samples provided in the multimedia files contain the full spectrum of resonances up to the imposed 6000 Hz cutoff frequency. The effect of vocal tract morphology on the third and fourth resonances will be important to understand as well, and may reveal sex and age dependencies that occur during development. The standing wave pattern for higher frequency resonances is more complex though, and may require modeling techniques that explore the acoustic sensitivity of vocal tract configurations (Story, 2006; Adachi *et al.*, 2007) and the relation of upper resonances to the voice source (Kitamura *et al.*, 2006). Additionally, there are developmental differences between males and females not only in vocal tract morphology but also other aspects of development that may affect the acoustic characteristics of speech.

The approach described in this study for transforming an adult vocal tract model to one that represents younger ages, both male and female, is similar to some previous modeling studies in that a statistical model was used to represent the tract configuration, and warping of the length axis was associated with cross-dimension scaling. It is different in several ways, however. For example, the VLAM approach (Ménard *et al.*, 2004; Oohashi *et al.*, 2017) relies on a single scale factor in the lower pharynx, another in the oral cavity, and an interpolation between them to scale the midsagittal cross-distances according to age (but not sex). In contrast, the scaling functions derived in this study from new measurements of anatomic landmarks are continuous, nonuniform functions extending from glottis to lips, and are dependent on both age *and* sex. They can be applied directly to the three components of the adult-based area function model to generate new versions of the model representative of males or females at given ages. These are provided in Tables III and IV for ages 0, 4, 8, and 12 yrs and adult, and can be easily used in combination with Eq. (1) and the coefficients in Table II or the coefficient grid in Fig. 1(b), to generate a wide range vocal tract area functions at each of the respective ages. Modeling the vocal tract shape directly at the level of the area function rather than the midsagittal plane is also an advantage because it eliminates the need to transform midsagittal cross-distances to areas (Heinz and Stevens, 1965; Mermelstein, 1973). In addition, even though all child-like and adult female versions of the model in this study were derived from the vocal tract structure of a specific adult male (Story *et al.*, 1996), alternative versions, or entirely

new "families," could be easily generated based on area functions measured for other talkers. For example, Story (2005b) reported sets of 11 vowel area functions of three adult males and three adult females, any of which could serve as the initial vocal tract model from which child-like versions are derived.

Another advantage is that, regardless of age or sex, the vocal tract model requires specification of only two parameters, the q_1 and q_2 coefficients in Eq. (1) [also Fig. 1(b) and Table III], to generate an area function for a vowel. Importantly, these two model parameters map in a one-to-one fashion to the first two resonance frequencies, or formants, of the vocal tract, and maintain this mapping across the age range examined here. An application of this relation is that it allows for mapping formant tracks extracted from an audio recording of a child's utterance to a sequence of vocal tract area functions indicating the time-dependence of the vocal tract configuration. An example is shown in Fig. 14(a) where an $[F_1, F_2]$ trajectory measured from a 4 year-old male talker's production of the target word "Iowa" (Bunton and Story, 2016) is superimposed on the vowel space generated by the four-year old male vocal tract model (i.e., from Fig. 13). Based on the proximity of points along the trajectory to those in the vowel space mesh, the corresponding trajectory in the $[q_1, q_2]$ coefficient space is determined [see Fig. 1(b)] and used in conjunction with Eq. (1) to produce a time-varying area function. The vocal tract shapes for the word Iowa spoken by the 4-year old male are shown in pseudo-midsagittal form in Fig. 14(b). By coupling this sequence to the same type of voice source used to generate the vowel samples, a simulation of the word can be produced. It is available for listening in the audio file **Mm. 3**.

Mm. 3 Audio file. This is a file of type mp3 (400 Kb).

The capability of relating formants measured from vowel-like utterances (i.e., vowels and vowel-to-vowel transitions) of a child at a given age and sex, to time-varying vocal tract shapes, as well as to simulations of the original utterances, provides new opportunities for investigating aspects of the production and perception of some types of speech sounds during development. Variability of formant trajectories, for example, could be related to vocal tract shape variability relative to age or sex. This may offer some insight into the timing of vocal tract movements relative to males versus females or across an age range. Some aspects of speech disorders could also be studied by either mapping formants extracted from vowel-like productions of disordered speech to vocal tract shapes, or by modifying simulations of typically developing children to become disordered. Simulations such as these might be used as stimuli for perceptual experiments in which listeners' responses could be related to systematic modifications of both the spatial and temporal aspects of the vocal tract configuration.

VIII. CONCLUSION

This study was a first step toward constructing a developmental and sex-specific version of a parametric area function model. A method was developed that utilizes CT-based

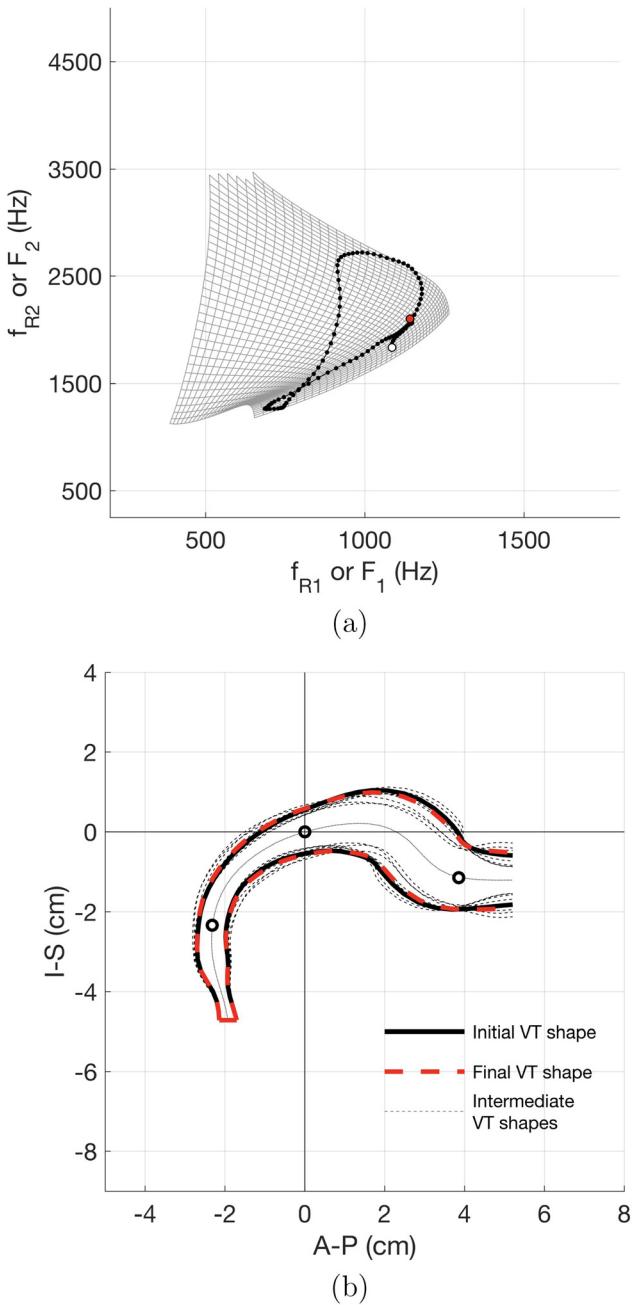


FIG. 14. (Color online) Demonstration of mapping formant trajectories of the target word Iowa extracted from an audio recording of a 4 yr-old male talker to a temporal sequence of vocal tract shapes. (a) Vowel space mesh generated with the 4 yr-old male vocal tract model along with the measured $[F_1, F_2]$ trajectory. The white dot denotes the beginning of the utterance and the larger solid dot (red) indicates the end. (b) Pseudo-midsagittal projections of time-dependent area functions that would produce Iowa. As in previous figures, the origin of the plot is located at the HPSP junction (landmark B).

measurements of age and sex specific anatomic landmarks within the vocal tract to transform an adult model to one that represents a hypothetical male or female child at a target age. Specifically, locations of glottis, superior point of the epiglottis, junction of the hard and soft palates, posterior or buccal aspect of the upper and lower lips, and the anterior aspect of the lips, form a 5-element vector that dictates how the length axis of the adult vocal tract must be warped to become child-like for given age and sex. The length warped vocal tract axis was then used to determine age-dependent cross-dimension scaling functions

that nonuniformly attenuate the components of the vocal tract model. The plausibility of vocal tract shapes generated by the age-transformed versions of the model was assessed by (1) comparing them to midsagittal images obtained from children in a quiet breathing condition, (2) applying the transformation to adult APh area functions and comparing the output to APh area functions measured for children, and (3) by comparing model-generated vowel space plots to measured formant data. These assessments indicated that the vocal tract model can be transformed to be reasonably child-like. The results of this study provide a tool that may be used to provide insights regarding the possibilities of *sound* production by small vocal tracts, which can facilitate understanding actual *vowel* production as well as studies of the perception of child-like vowel utterances.

ACKNOWLEDGMENTS

Research supported by Grant Nos. NIH R01-DC006282, NIH R01-DC011275, NSF BCS-1145011, and NIH P30 HD03352. The authors express their gratitude to the late Ralph Ohde for the age and vowel specific data from the [Perry et al. \(2001\)](#) publication. H.K.V. gratefully acknowledges Dr. Lindell Gentry for his assistance with establishing the Vocal Tract Development Lab's lifespan imaging database, and the following individuals for assistance in placing anatomic landmarks for the development of sex-specific vocal tract growth models: Celia Choih, Michael Kelly, and Ying Ji Chuang; also, special thanks to Shannon Theis, Erin Douglas, Carlyn Burris, Sara Kurtzweil, and Katelyn Kassulke Tillman for assistance in establishing the Acoustic Pharyngometry protocol for data collection and data analysis.

APPENDIX

The components of the vocal tract model represented by Eq. (1) are provided in Tables II, III, and IV. The q_1 and q_2 coefficients that correspond to the original ten area functions on which the model is based are in Table II. The $\Omega(i)$, $\phi_1(i)$, and $\phi_2(i)$ for the original adult male and derived male versions are given in Table III. Element 1 is located just above the glottis and element 44 is the lip termination. The bottom two rows in the table indicate the element length $L(i)$ and the total VTL for each age. The derived female versions of the model are arranged in the same manner in Table IV.

TABLE II. Mode coefficients, q_1 and q_2 , that can be used to reconstruct the vowels indicated in the first column based on Eq. (1).

Vowel	q_1	q_2
i	-5.10	0.88
ɪ	-2.57	0.93
ɛ	-1.07	1.22
æ	0.66	2.22
ʌ	2.55	0.10
ɑ	3.86	1.35
ɔ	3.47	-0.45
ɒ	0.00	-2.69
ʊ	1.68	-1.87
ʊ	-3.48	-1.70

TABLE III. Numerical form of the $\Omega(i)$, $\phi_1(i)$, and $\phi_2(i)$ model components represented by Eq. (1) and Fig. 1, for the original adult male (shown with the asterisk) and derived male versions of the model at ages 0, 4, 8, and 12 yrs. Element 1 is located just above the glottis and element 44 is the lip termination. The bottom two rows in the table indicate the element length $L(i)$ and the total VTL for each age.

Element i	Adult male*			12 yr male			8 yr male			4 yr male			0 yr male		
	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$												
1	0.80	0.004	0.003	0.58	0.003	0.002	0.50	0.003	0.002	0.43	0.002	0.002	0.34	0.002	0.001
2	0.74	-0.014	-0.020	0.53	-0.011	-0.017	0.46	-0.010	-0.016	0.40	-0.009	-0.014	0.31	-0.007	-0.010
3	0.71	-0.029	-0.052	0.52	-0.024	-0.044	0.45	-0.021	-0.040	0.39	-0.019	-0.036	0.31	-0.014	-0.026
4	0.72	-0.042	-0.084	0.55	-0.035	-0.069	0.49	-0.031	-0.062	0.43	-0.027	-0.055	0.32	-0.020	-0.040
5	0.85	-0.054	-0.109	0.71	-0.044	-0.087	0.65	-0.040	-0.077	0.58	-0.035	-0.068	0.41	-0.026	-0.051
6	1.10	-0.066	-0.124	0.94	-0.054	-0.096	0.85	-0.049	-0.084	0.75	-0.043	-0.073	0.55	-0.032	-0.056
7	1.35	-0.078	-0.131	1.08	-0.065	-0.097	0.95	-0.058	-0.084	0.82	-0.052	-0.072	0.63	-0.038	-0.057
8	1.47	-0.091	-0.129	1.09	-0.077	-0.092	0.93	-0.069	-0.078	0.80	-0.062	-0.066	0.63	-0.044	-0.054
9	1.44	-0.105	-0.121	1.05	-0.090	-0.083	0.91	-0.081	-0.069	0.79	-0.073	-0.057	0.61	-0.052	-0.048
10	1.39	-0.120	-0.108	1.05	-0.104	-0.070	0.92	-0.094	-0.056	0.80	-0.084	-0.046	0.61	-0.060	-0.041
11	1.38	-0.137	-0.092	1.08	-0.118	-0.055	0.95	-0.107	-0.043	0.84	-0.096	-0.033	0.63	-0.068	-0.032
12	1.42	-0.153	-0.073	1.12	-0.132	-0.038	0.99	-0.119	-0.027	0.86	-0.106	-0.020	0.65	-0.076	-0.022
13	1.46	-0.170	-0.053	1.15	-0.144	-0.021	1.00	-0.129	-0.011	0.87	-0.114	-0.005	0.67	-0.083	-0.012
14	1.49	-0.184	-0.032	1.15	-0.153	-0.002	1.00	-0.136	0.006	0.87	-0.119	0.011	0.67	-0.089	-0.001
15	1.49	-0.195	-0.009	1.16	-0.158	0.018	1.02	-0.139	0.024	0.88	-0.120	0.028	0.67	-0.092	0.011
16	1.49	-0.203	0.015	1.18	-0.159	0.039	1.04	-0.137	0.044	0.91	-0.117	0.046	0.68	-0.092	0.023
17	1.50	-0.205	0.040	1.21	-0.155	0.062	1.07	-0.131	0.064	0.93	-0.108	0.064	0.70	-0.089	0.036
18	1.53	-0.201	0.067	1.24	-0.144	0.085	1.09	-0.119	0.085	0.94	-0.095	0.082	0.72	-0.083	0.049
19	1.57	-0.191	0.095	1.26	-0.128	0.108	1.10	-0.102	0.106	0.94	-0.078	0.100	0.72	-0.073	0.063
20	1.59	-0.174	0.123	1.25	-0.106	0.131	1.10	-0.079	0.125	0.95	-0.056	0.116	0.72	-0.060	0.076
21	1.58	-0.149	0.150	1.25	-0.079	0.151	1.10	-0.053	0.142	0.96	-0.031	0.130	0.72	-0.044	0.088
22	1.57	-0.118	0.176	1.26	-0.047	0.168	1.11	-0.023	0.155	0.96	-0.003	0.139	0.72	-0.025	0.097
23	1.56	-0.082	0.198	1.26	-0.012	0.180	1.10	0.010	0.161	0.95	0.026	0.141	0.72	-0.004	0.103
24	1.56	-0.040	0.214	1.24	0.026	0.184	1.08	0.044	0.163	0.93	0.056	0.139	0.70	0.018	0.105
25	1.53	0.005	0.224	1.21	0.064	0.182	1.07	0.077	0.156	0.95	0.085	0.130	0.69	0.040	0.102
26	1.50	0.051	0.224	1.21	0.101	0.170	1.09	0.109	0.141	0.99	0.112	0.114	0.69	0.062	0.094
27	1.46	0.097	0.214	1.24	0.136	0.149	1.15	0.139	0.119	1.06	0.137	0.090	0.71	0.081	0.079
28	1.46	0.141	0.193	1.32	0.167	0.120	1.23	0.164	0.088	1.13	0.157	0.061	0.74	0.097	0.059
29	1.52	0.180	0.160	1.40	0.193	0.082	1.30	0.185	0.052	1.21	0.174	0.028	0.78	0.110	0.034
30	1.60	0.213	0.117	1.48	0.212	0.039	1.38	0.199	0.012	1.28	0.184	-0.009	0.82	0.118	0.007
31	1.68	0.238	0.066	1.56	0.224	-0.008	1.46	0.207	-0.030	1.37	0.189	-0.045	0.87	0.122	-0.020
32	1.76	0.255	0.009	1.66	0.229	-0.054	1.56	0.208	-0.070	1.46	0.188	-0.078	0.92	0.121	-0.045
33	1.86	0.261	-0.048	1.76	0.226	-0.095	1.65	0.202	-0.103	1.54	0.181	-0.105	0.98	0.116	-0.066
34	1.97	0.257	-0.100	1.84	0.215	-0.128	1.72	0.191	-0.128	1.60	0.169	-0.122	1.03	0.108	-0.078
35	2.06	0.244	-0.140	1.91	0.198	-0.146	1.78	0.173	-0.134	1.65	0.153	-0.123	1.07	0.097	-0.079
36	2.13	0.221	-0.163	1.95	0.175	-0.145	1.81	0.152	-0.130	1.66	0.133	-0.113	1.08	0.084	-0.070
37	2.17	0.192	-0.160	1.94	0.148	-0.123	1.79	0.128	-0.102	1.63	0.111	-0.083	1.07	0.070	-0.047
38	2.13	0.158	-0.129	1.86	0.120	-0.077	1.70	0.103	-0.053	1.55	0.089	-0.036	1.03	0.056	-0.010
39	2.01	0.122	-0.065	1.74	0.092	-0.008	1.59	0.079	0.014	1.46	0.069	0.028	0.99	0.043	0.038
40	1.84	0.089	0.028	1.63	0.067	0.080	1.50	0.058	0.097	1.38	0.051	0.103	0.98	0.034	0.094
41	1.69	0.061	0.141	1.53	0.049	0.178	1.42	0.043	0.185	1.31	0.038	0.183	0.97	0.027	0.153
42	1.55	0.043	0.256	1.44	0.038	0.271	1.35	0.036	0.269	1.25	0.033	0.257	0.96	0.025	0.210
43	1.43	0.038	0.344	1.35	0.036	0.331	1.28	0.035	0.317	1.19	0.033	0.297	0.95	0.026	0.238
44	1.31	0.047	0.356	1.26	0.045	0.342	1.21	0.044	0.328	1.13	0.041	0.307	0.92	0.033	0.248
L(i) (cm)	0.400			0.330			0.296			0.264			0.192		
VTL (cm)	17.6			14.52			13.02			11.62			8.45		

TABLE IV. Numerical form of the $\Omega(i)$, $\phi_1(i)$, and $\phi_2(i)$ model components represented by Eq. (1) for derived female versions of the model at ages 0, 4, 8, 12 yrs and adult. Element 1 is located just above the glottis and element 44 is the lip termination. The bottom two rows in the table indicate the element length $L(i)$ and the total VTL for each age.

Element i	Adult female			12 yr female			8 yr female			4 yr female			0 yr female		
	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$
1	0.63	0.003	0.002	0.53	0.003	0.002	0.44	0.002	0.002	0.38	0.002	0.001	0.37	0.002	0.001
2	0.58	-0.012	-0.018	0.49	-0.011	-0.016	0.40	-0.010	-0.016	0.35	-0.009	-0.015	0.34	-0.006	-0.009

TABLE IV. (Continued)

Element i	Adult female			12 yr female			8 yr female			4 yr female			0 yr female		
	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$
3	0.56	-0.026	-0.047	0.48	-0.023	-0.043	0.40	-0.021	-0.040	0.35	-0.019	-0.036	0.33	-0.014	-0.025
4	0.60	-0.037	-0.074	0.52	-0.033	-0.066	0.47	-0.030	-0.060	0.42	-0.027	-0.054	0.33	-0.020	-0.040
5	0.76	-0.048	-0.093	0.69	-0.042	-0.082	0.65	-0.039	-0.072	0.59	-0.035	-0.064	0.40	-0.025	-0.051
6	1.00	-0.058	-0.104	0.90	-0.052	-0.090	0.82	-0.048	-0.077	0.73	-0.043	-0.067	0.52	-0.031	-0.057
7	1.17	-0.069	-0.105	1.02	-0.062	-0.090	0.87	-0.058	-0.076	0.75	-0.052	-0.065	0.62	-0.036	-0.058
8	1.18	-0.082	-0.101	1.01	-0.074	-0.085	0.85	-0.070	-0.069	0.73	-0.063	-0.058	0.64	-0.042	-0.056
9	1.13	-0.096	-0.091	0.98	-0.086	-0.075	0.85	-0.082	-0.059	0.74	-0.074	-0.048	0.61	-0.049	-0.050
10	1.14	-0.111	-0.077	0.99	-0.100	-0.063	0.88	-0.095	-0.047	0.77	-0.086	-0.037	0.60	-0.056	-0.043
11	1.17	-0.126	-0.061	1.03	-0.114	-0.048	0.93	-0.108	-0.033	0.81	-0.097	-0.024	0.60	-0.063	-0.034
12	1.21	-0.141	-0.044	1.06	-0.126	-0.033	0.96	-0.119	-0.019	0.83	-0.106	-0.011	0.61	-0.070	-0.023
13	1.24	-0.154	-0.025	1.08	-0.138	-0.015	0.98	-0.129	-0.003	0.84	-0.113	0.004	0.61	-0.075	-0.012
14	1.25	-0.164	-0.005	1.09	-0.146	0.003	0.99	-0.135	0.014	0.86	-0.117	0.019	0.59	-0.078	-0.001
15	1.25	-0.170	0.016	1.10	-0.150	0.022	1.02	-0.138	0.032	0.88	-0.117	0.036	0.58	-0.079	0.011
16	1.27	-0.172	0.039	1.12	-0.150	0.043	1.06	-0.136	0.052	0.92	-0.113	0.054	0.58	-0.076	0.024
17	1.30	-0.168	0.063	1.15	-0.144	0.064	1.10	-0.130	0.072	0.94	-0.103	0.072	0.58	-0.070	0.037
18	1.34	-0.158	0.088	1.18	-0.133	0.086	1.12	-0.118	0.092	0.95	-0.090	0.090	0.59	-0.060	0.050
19	1.35	-0.141	0.113	1.19	-0.116	0.109	1.13	-0.100	0.113	0.96	-0.072	0.107	0.58	-0.047	0.062
20	1.35	-0.119	0.137	1.19	-0.094	0.130	1.13	-0.078	0.131	0.97	-0.050	0.123	0.58	-0.031	0.073
21	1.35	-0.090	0.159	1.19	-0.067	0.149	1.14	-0.052	0.147	0.97	-0.024	0.135	0.59	-0.013	0.082
22	1.35	-0.057	0.178	1.20	-0.036	0.164	1.14	-0.023	0.159	0.98	0.003	0.142	0.59	0.007	0.087
23	1.35	-0.019	0.192	1.20	-0.002	0.174	1.13	0.009	0.166	0.97	0.032	0.145	0.59	0.027	0.089
24	1.34	0.020	0.197	1.18	0.034	0.176	1.11	0.042	0.167	0.95	0.061	0.141	0.60	0.047	0.086
25	1.31	0.061	0.197	1.16	0.070	0.173	1.09	0.075	0.160	0.97	0.089	0.131	0.63	0.067	0.078
26	1.30	0.101	0.186	1.17	0.105	0.160	1.10	0.107	0.146	1.01	0.115	0.115	0.69	0.086	0.067
27	1.32	0.139	0.166	1.22	0.138	0.140	1.16	0.136	0.125	1.09	0.140	0.094	0.76	0.103	0.051
28	1.40	0.174	0.136	1.30	0.167	0.112	1.23	0.161	0.097	1.17	0.161	0.067	0.83	0.118	0.031
29	1.49	0.203	0.097	1.39	0.192	0.076	1.31	0.183	0.063	1.25	0.179	0.035	0.90	0.131	0.007
30	1.58	0.225	0.052	1.47	0.211	0.035	1.38	0.199	0.025	1.33	0.192	0.001	0.98	0.139	-0.019
31	1.66	0.240	0.002	1.56	0.223	-0.009	1.46	0.209	-0.016	1.42	0.199	-0.035	1.06	0.143	-0.044
32	1.76	0.246	-0.048	1.65	0.228	-0.053	1.55	0.212	-0.055	1.51	0.200	-0.070	1.14	0.143	-0.067
33	1.87	0.244	-0.094	1.75	0.226	-0.093	1.64	0.209	-0.091	1.59	0.196	-0.099	1.20	0.138	-0.085
34	1.96	0.234	-0.131	1.83	0.215	-0.124	1.70	0.198	-0.118	1.65	0.185	-0.121	1.25	0.129	-0.095
35	2.03	0.216	-0.155	1.88	0.198	-0.144	1.75	0.182	-0.134	1.69	0.169	-0.128	1.28	0.117	-0.095
36	2.08	0.192	-0.155	1.92	0.176	-0.143	1.78	0.161	-0.132	1.70	0.149	-0.126	1.28	0.102	-0.087
37	2.07	0.163	-0.138	1.90	0.150	-0.126	1.76	0.137	-0.115	1.67	0.126	-0.104	1.25	0.086	-0.064
38	2.00	0.132	-0.093	1.83	0.121	-0.086	1.69	0.111	-0.076	1.58	0.102	-0.066	1.18	0.069	-0.029
39	1.86	0.102	-0.023	1.69	0.093	-0.023	1.56	0.085	-0.018	1.46	0.078	-0.011	1.10	0.053	0.017
40	1.72	0.074	0.069	1.55	0.068	0.058	1.43	0.062	0.056	1.33	0.056	0.057	1.02	0.039	0.072
41	1.60	0.053	0.172	1.43	0.048	0.149	1.32	0.044	0.140	1.21	0.040	0.132	0.96	0.028	0.129
42	1.49	0.040	0.272	1.32	0.036	0.237	1.22	0.033	0.220	1.11	0.030	0.203	0.90	0.024	0.182
43	1.39	0.037	0.340	1.22	0.032	0.298	1.13	0.030	0.275	1.02	0.027	0.250	0.85	0.023	0.211
44	1.29	0.047	0.351	1.13	0.041	0.306	1.04	0.038	0.282	0.94	0.034	0.256	0.80	0.029	0.217
$L(i)$ (cm)	0.353			0.315			0.288			0.259			0.192		
VTL (cm)	15.53			13.86			12.67			11.4			8.45		

¹According to the conventions recently proposed by Titze *et al.* (2015), the vocal tract resonance frequencies determined from a direct calculation of the frequency response are denoted as f_{Rn} , whereas formant frequencies measured from the acoustic signal by processing algorithms are denoted as F_n .
²To obtain profiles representative across age, the 44-point centerlines shown previously in Fig. 2, both male and female, were first averaged, point by point, within each yearly increment (i.e., 0–1 yrs, 1–2 yrs, etc.). Each point along each mean centerline was then fit with a third-order polynomial across age, resulting in 44 separate polynomials that describe the changing shape of the vocal tract profile. This allows for reconstructing a profile at any target age within the age continuum.

Adachi, S., Takemoto, H., Kitamura, T., Mokhtari, P., and Honda, K. (2007). “Vocal tract length perturbation and its application to male-female vocal tract shape conversion,” *J. Acoust. Soc. Am.* **121**(6), 3874–3885.

- Alwan, A. A., Narayanan, S. S., and Haker, K. (1997). “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics,” *J. Acoust. Soc. Am.* **101**(2), 1078–1089.
 Assmann, P. F., and Katz, W. F. (2000). “Time-varying spectral change in the vowels of children and adults,” *J. Acoust. Soc. Am.* **108**(4), 1856–1866.
 Badin, P., Boë, L. J., Sawallis, T. R., and Schwartz, J. L. (2014). “Keep the lips to free the larynx: Comments on de Boer’s articulatory model (2010),” *J. Phonetics* **46**, 161–167.
 Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (1991). “Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,” *J. Acoust. Soc. Am.* **90**, 799–828.
 Barbier, G., Boë, L. J., Captier, G., and Laboissiere, R. (2015). “Human vocal tract growth: A longitudinal study of the development of various anatomical structures,” in *Proceedings of the 16th Annual Conference on International Speech Communication Association (Interspeech 2015)*.

- Beautemps, D., Badin, P., and Laboissiere, R. (1995). "Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data," *Speech Commun.* **19**(16), 27–47.
- Boë, L. J. (1999). "Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: Consequences for ontogenesis and phylogenesis," in *Proceedings of the International Congress of Phonetic Sciences*, Vol. 3, No. 25, pp. 1–25.
- Boë, L. J., Badin, P., Ménard, L., Captier, G., Davis, B., MacNeilage, P., and Schwartz, J. L. (2013). "Anatomy and control of the developing human vocal tract: A response to Lieberman," *J. Phonetics* **41**(5), 379–392.
- Boë, L. J., Granat, J., Badin, P., Autesserre, D., Pochic, D., Zga, N., and Ménard, L. (2006). "Skull and vocal tract growth from newborn to adult," in *Proceedings of the 7th International Seminar on Speech Production, ISSP7*, UFMG, Belo Horizonte, Brazil, pp. 75–82.
- Boë, L. J., Heim, J. L., Honda, K., Maeda, S., Badin, P., and Abry, C. (2007). "The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a)," *J. Phonetics* **35**(4), 564–581.
- Boë, L. J., and Maeda, S. (1998). "Modélisation de la croissance du conduit vocal" ("Modeling vocal tract growth"), *Journées d'Études Linguistiques, La Voyelle dans tous ses états* (Journal of Linguistic Studies, The vowel in all its states), 98–105.
- Buhr, R. D. (1980). "The emergence of vowels in an infant," *J. Speech, Lang., Hear. Res.* **23**(1), 73–94.
- Bunton, K., and Story, B. H. (2016). "Arizona Child Acoustic Database Repository," *Folia Phoniatrica et Logopaedica* **68**(3), 107–111.
- Bunton, K., Story, B. H., and Titze, I. (2013). "Estimation of vocal tract area functions in children based on measurement of lip termination area and inverse acoustic mapping," *Proc. Mtgs. Acoust.* **19**(1), 060054.
- Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (2000). "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," *J. Speech, Lang., Hear. Res.* **43**(3), 721–736.
- Dang, J., and Honda, K. (1997). "Acoustic characteristics of the piriform fossa in models and humans," *J. Acoust. Soc. Am.* **101**(1), 456–465.
- de Boer, B. (2010). "Modelling vocal anatomy's significant effect on speech," *J. Evol. Psych.* **8**(4), 351–366.
- de Boer, B., and Fitch, W. T. (2010). "Computer models of vocal tract evolution: An overview and critique," *Adap. Behav.* **18**(1), 36–47.
- Delvaux, B., and Howard, D. (2014). "A new method to explore the spectral impact of the piriform fossae on the singing voice: Benchmarking using MRI-based 3D-printed vocal tracts," *PLoS One* **9**(7), e102680.
- Eguchi, S., and Hirsh, I. J. (1969). "Development of speech sounds in children," *Acta Oto-laryngologica Suppl.* **257**, 1–51.
- Fant, G. (1960). *The Acoustic Theory of Speech Production* (Mouton, The Hague).
- Fant, G. (1966). "A note on vocal tract size factors and non-uniform F-pattern scalings," *Speech Transmiss. Lab. Q. Prog. Status Rep.* **1**, 22–30.
- Fant, G. (1975). "Non-uniform vowel normalization," *Speech Transmiss. Lab. Q. Progress Status Rep.* **16**(2–3), 1–19.
- 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.
- Fritsch, F. N., and Carlson, R. E. (1980). "Monotone piecewise cubic interpolation," *SIAM J. Numer. Anal.* **17**, 238–246.
- Gilbert, H. R., Robb, M. P., and Chen, Y. (1997). "Formant frequency development: 15 to 36 months," *J. Voice* **11**(3), 260–266.
- Goldstein, U. G. (1980). "An articulatory model for the vocal tracts of growing children," Doctoral dissertation, Department of Electrical Engineering and Computer Science, MIT.
- Guenther, F. H., Hampson, M., and Johnson, D. (1998). "A theoretical investigation of reference frames for the planning of speech movements," *Psych. Rev.* **105**(4), 611–633.
- Heinz, J. M., and Stevens, K. N. (1965). "On the relations between lateral cineradiographs, area functions and acoustic spectra of speech," in *Proceedings of the 5th Congres International D'Acoustique*.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**(5), 3099–3111.
- Honda, K., Kitamura, T., Takemoto, H., Adachi, S., Mokhtari, P., Takano, S., and Masaki, S. (2010). "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," *Comp. Methods Biomed. Eng.* **13**(4), 443–453.
- 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," *Archives Oral Biol.* **77**, 27–38.
- Kent, R. D., and Murray, A. D. (1982). "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," *J. Acoust. Soc. Am.* **72**(2), 353–365.
- Kitamura, T., Takemoto, H., Adachi, S., Mokhtari, P., and Honda, K. (2006). "Cyclicity of laryngeal cavity resonance due to vocal fold vibration," *J. Acoust. Soc. Am.* **120**(4), 2239–2249.
- 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.
- Lieberman, D. E., and McCarthy, R. C. (1999). "The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions," *J. Human Evol.* **36**(5), 487–517.
- Lieberman, D. E., McCarthy, R. C., Hiemae, K. M., and Palmer, J. B. (2001). "Ontogeny of postnatal hyoid and larynx descent in humans," *Archives Oral Biol.* **46**(2), 117–128.
- Lieberman, P. (2006). "Limits on tongue deformation—Diana monkey formants and the impossible vocal tract shapes proposed by Riede *et al.* (2005)," *J. Hum. Evol.* **50**(2), 219–221.
- Lieberman, P. (2007). "Current views on Neanderthal speech capabilities: A reply to Boe *et al.* (2002)," *J. Phonetics* **35**(4), 552–563.
- Lieberman, P. (2012). "Vocal tract anatomy and the neural bases of talking," *J. Phonetics* **40**(4), 608–622.
- Lieberman, P., and Crelin, E. S. (1971). "On the speech of Neanderthal man," *Ling. Inq.* **2**(2), 203–222.
- Lieberman, P., Crelin, E. S., and Klatt, D. H. (1972). "Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee," *Am. Anthropol.* **74**(3), 287–307.
- Maeda, S. (1979). "An articulatory model of the tongue based on a statistical analysis," *J. Acoust. Soc. Am.* **65**, S22 (1979).
- Maeda, S. (1990). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modelling*, edited by W. L. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, The Netherlands), pp. 131–149.
- Mathworks (2016). MATLAB, Version 9.0.0.34160, (R2016a).
- McGowan, R. S., Jackson, M. T., and Berger, M. A. (2012). "Analyses of vocal tract cross-distance to area mapping: An investigation of a set of vowel images," *J. Acoust. Soc. Am.* **131**(1), 424–434 (2012).
- Ménard, L., Davis, B. L., Boë, L. J., and Roy, J. P. (2009). "Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels," *J. Speech, Lang., Hear. Res.* **52**(5), 1268–1285.
- Ménard, L., Schwartz, J. L., and Boë, L. J. (2004). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood," *J. Speech, Lang., Hear. Res.* **47**(5), 1059–1080.
- Mermelstein, P. (1973). "Articulatory model for the study of speech production," *J. Acoust. Soc. Am.* **53**(4), 1070–1082.
- Mokhtari, P., Kitamura, T., Takemoto, H., and Honda, K. (2007). "Principal components of vocal tract area functions and inversion of vowels by linear regression of cepstrum coefficients," *J. Phonetics* **35**, 20–39.
- Narayanan, S. S., Alwan, A. A., and Haker, K. (1995). "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Am.* **98**(3), 1325–1347.
- Narayanan, S. S., Alwan, A. A., and Haker, K. (1997a). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," *J. Acoust. Soc. Am.* **101**(2), 1064–1077.
- Narayanan, S. S., Alwan, A. A., and Song, Y. (1997b). "New results in vowel production: MRI, EPG, and acoustic data," in *Proceedings of the 5th European Conference on Speech Communication and Technology, EUROSPEECH 1997*, Rhodes, Greece (September 22–25), Vol. 2, pp. 1007–1009.
- Nordström, P.-E. (1977). "Female and infant vocal tracts simulated from male area functions," *J. Phonetics* **5**, 81–92.
- Oohashi, H., Watanabe, H., and Taga, G. (2017). "Acquisition of vowel articulation in childhood investigated by acoustic-to-articulatory inversion," *Inf. Behav. Devel.* **46**, 178–193.
- 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.
- Pruthi, T., Espy-Wilson, C. Y., and Story, B. H. (2007). "Simulation and analysis of nasalized vowels based on magnetic resonance imaging data," *J. Acoust. Soc. Am.* **121**(6), 3858–3873.
- Riely, R. R., and Smith, A. (2003). "Speech movements do not scale by orofacial structure size," *J. Appl. Physiol.* **94**(6), 2119–2126.
- Sondhi, M. M., and Schroeter, J. (1987). "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. ASSP ASSP-35*(7), 955–967.
- Stevens, K. N. (2000). *Acoustic Phonetics* (MIT Press, Cambridge, MA), Vol. 30.
- Story, B. H. (2005a). "A parametric model of the vocal tract area function for vowel and consonant simulation," *J. Acoust. Soc. Am.* **117**(5), 3231–3254.
- Story, B. H. (2005b). "Synergistic modes of vocal tract articulation for American English vowels," *J. Acoust. Soc. Am.* **118**(6), 3834–3859.
- Story, B. H. (2006). "Technique for tuning vocal tract area functions based on acoustic sensitivity functions," *J. Acoust. Soc. Am.* **119**(2), 715–718.
- Story, B. H. (2009). "Vocal tract modes based on multiple area function sets from one speaker," *J. Acoust. Soc. Am.* **125**(4), EL141–EL147.
- Story, B. H. (2013). "Phrase-level speech simulation with an airway modulation model of speech production," *Comp. Speech Lang.* **27**(4), 989–10105.
- Story, B. H., Laukkanen, A.-M., and Titze, I. R. (2000). "Acoustic impedance of an artificially lengthened and constricted vocal tract," *J. Voice* **14**(4), 455–469.
- Story, B. H., and Titze, I. R. (1998). "Parameterization of vocal tract area functions by empirical orthogonal modes," *J. Phonetics* **26**(3), 223–260.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.* **100**(1), 537–554.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (2001). "The relationship of vocal tract shape to three voice qualities," *J. Acoust. Soc. Am.* **109**(4), 1651–1667.
- Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. (2006). "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *J. Acoust. Soc. Am.* **119**(2), 1037–1049.
- Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Kreiman, J., Kob, M., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., and Wolfe, J. (2015). "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," *J. Acoust. Soc. Am.* **137**(5), 3005–3007.
- Titze, I. R., Horii, Y., and Scherer, R. C. (1987). "Some technical considerations in voice perturbation measurements," *J. Speech. Hear. Res.* **30**, 252–260.
- Vampola, T., Horacek, J., and Švec, J. G. (2015). "Modeling the influence of piriform sinuses and valleculae on the vocal tract resonances and anti-resonances," *Acta Acust. Acust.* **101**(3), 594–602.
- 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**, 1510–1545.
- Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (1999). "Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results," *Int. J. Ped. Otorhinolaryngol.* **49**(3), 197–206.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (2005). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," *J. Acoust. Soc. Am.* **117**(1), 338–350.
- Vorperian, H. K., Kurtzweil, S. L., Fourakis, M., Kent, R. D., Tillman, K. K., and Austin, D. (2015). "Effect of body position on vocal tract acoustics: Acoustic pharyngometry and vowel formants," *J. Acoust. Soc. Am.* **138**(2), 833–845.
- 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. B., Ziegert, A. J., 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.
- Yang, C.-S., and Kasuya, H. (1994). "Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects," in *Proceedings of ICSLP 94*, Yokohama, Japan, pp. 623–626.