

Speech production variability in fricatives of children and adults: Results of functional data analysis

Laura L. Koenig^{a)}

Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511 and Long Island University,
One University Plaza, Brooklyn, New York 11201

Jorge C. Lucero^{b)}

Department of Mathematics, University of Brasilia, Brasilia DF 70910-900 Brazil

Elizabeth Perlman^{c)}

Rusk Institute, New York University Medical Center, 400 East 34th Street, RM 517, New York, New York
10016

(Received 3 February 2008; revised 29 July 2008; accepted 13 August 2008)

This study investigates token-to-token variability in fricative production of 5 year olds, 10 year olds, and adults. Previous studies have reported higher intrasubject variability in children than adults, in speech as well as nonspeech tasks, but authors have disagreed on the causes and implications of this finding. The current work assessed the characteristics of age-related variability across articulators (larynx and tongue) as well as in temporal versus spatial domains. Oral airflow signals, which reflect changes in both laryngeal and supralaryngeal apertures, were obtained for multiple productions of /h s z/. The data were processed using functional data analysis, which provides a means of obtaining relatively independent indices of amplitude and temporal (phasing) variability. Consistent with past work, both temporal and amplitude variabilities were higher in children than adults, but the temporal indices were generally less adultlike than the amplitude indices for both groups of children. Quantitative and qualitative analyses showed considerable speaker- and consonant-specific patterns of variability. The data indicate that variability in /s/ may represent laryngeal as well as supralaryngeal control and further that a simple random noise factor, higher in children than in adults, is insufficient to explain developmental differences in speech production variability. © 2008 Acoustical Society of America. [DOI: 10.1121/1.2981639]

PACS number(s): 43.70.Ep, 43.70.Aj [BHS]

Pages: 3158–3170

I. INTRODUCTION

Many past studies have demonstrated decreasing token-to-token variability in speech and other motor acts with increasing age, but interpretations of this finding have varied widely (see Secs. I B and I C). Some authors have argued that immature physiology yields more random sensorimotor noise in children than in adults, whereas others have emphasized patterned variation reflecting individual learning and exploration in the context of ongoing anatomical or physiological change. This study explored the characteristics of speech production variability in children using two strategies. First, variability was compared across two articulator systems, the larynx and the tongue, by analyzing oral airflow data in the fricatives /h s z/. This methodology yields information on the timing and magnitude of laryngeal and lingual actions in a noninvasive fashion. Second, variability in repeated productions of these consonants was quantified using functional data analysis (FDA). FDA is a useful tool for this purpose because it allows calculation of relatively independent measures of temporal and amplitude variabilities. These

results may provide a more detailed understanding of how speech motor control develops across articulatory systems and in temporal as well as spatial domains. A fuller description of developmental variability may, in turn, speak to the underlying sources of this phenomenon.

The following sections review (a) articulatory, aerodynamic, and developmental characteristics of fricatives; (b) variability in children's speech and other motor behaviors, considering (c) issues of methodology and interpretation; and (d) the characteristics of FDA and its appropriateness for the current work.

A. Fricative production and development

The focus on /h s z/ in this work was motivated by two considerations. Most fundamentally, these three fricatives, in combination, provide insight into laryngeal, oral, and combined laryngeal-oral activities: /h/ requires vocal fold abduction but no particular supralaryngeal movements; /z/ requires a tongue-tip constriction, while the vocal folds remain adducted for voicing; and /s/ requires both laryngeal abduction for devoicing and a tongue-tip constriction. A useful feature of oral airflow data is that they permit noninvasive assessment of both articulatory actions: Airflow rises with vocal fold abduction and decreases with tongue-tip constrictions.

^{a)}Author to whom correspondence should be addressed; electronic mail: koenig@haskins.yale.edu

^{b)}Electronic mail: lucero@mat.unb.br

^{c)}Electronic mail: elizabeth.perlman@nyumc.org

In the case of /s/, the signals typically show an increase related to abduction and a superimposed valley reflecting lingual actions (Klatt *et al.*, 1968).

Further, from a developmental perspective, the fricatives /s z/ are of interest because they are acquired rather late, on average, for many children (Sander, 1972). Kent (1992) proposed that lingual fricatives are late-appearing sounds because they require fine control of tongue position and force. Yet fricatives also have specific aerodynamic requirements; in particular, they require sufficient airflow to generate turbulent noise (Howe and McGowan, 2005, Scully *et al.*, 1992, and Shadle, 1990). Indeed, direct laryngeal data show that adult speakers may use a slightly more open glottal configuration in /z/ compared to surrounding vowels, presumably to increase airflow a bit without actively inhibiting voicing (Lisker *et al.*, 1969; Sawashima, 1970). If lingual fricatives are difficult for children because of their supraglottal requirements, one would expect to see parallel patterns for /s/ and /z/ within individuals. On the other hand, the combination of a precise supraglottal constriction and abduction for devoicing and high airflow in /s/ may yield an added level of complexity compared to /z/.

There has been surprisingly little discussion of laryngeal function in /s/ acquisition. This is in marked contrast to the long history of studying children's laryngeal control for stop consonant voicing, especially via measures of voice onset time (VOT) [e.g., see Kewley-Port and Preston (1974) and Macken and Barton (1980)]. A consistent finding in the VOT literature has been that adultlike levels of variability are reached later for voiceless aspirated stops than for their unaspirated counterparts. The traditional explanation for this has been that voiceless aspirated stops are challenging for children because they require precise temporal coordination of vocal fold abduction and oral release (Kewley-Port and Preston, 1974). The data obtained by Koenig (2000) suggested an alternative explanation. That study compared the duration of devoiced regions in /h/ with VOTs in English aspirated /p t/ produced by 5 year olds and adults, and also assessed the degree of abduction in /h/ using oral airflow signals. The results showed significant correlations between devoiced regions in /h/ and /p t/ VOTs, as well as great variability in the extent and duration of abduction in /h/ for many children. If children are highly variable in their abduction patterns for /h/—a consonant that requires *only* a laryngeal gesture—then it is likely that their laryngeal control will also be variable for consonants that require abduction along with supraglottal articulations. The current work extends that logic to the study of voiceless fricatives. Specifically, comparing variability patterns in /s/ to those of /h/ and /z/ can help elucidate how /s/ reflects laryngeal as well as oral actions and, more generally, shed light on the degree to which developmental variability is articulator specific, both within and across individuals.

B. Variability in development

Children, as a group, demonstrate greater token-to-token variability than adults when repeating the same speech sequence for acoustic, aerodynamic,¹ electropalatographic, and

kinematic measures [e.g., see Cheng *et al.* (2007), Chermak and Schneiderman (1986), Eguchi and Hirsh (1969), Kent (1976), Kent and Forner (1980), Munson (2004), Ohde (1985), Sharkey and Folkins (1985), Smith and Goffman (1998), Smith and McLean-Muse (1986), Tingley and Allen (1975), and Zajac and Hackett (2002)]. Child-adult differences can persist well into the school-age years (Eguchi and Hirsh, 1969; Kent and Forner, 1980; Lee *et al.* 1999; Munson, 2004; Ohde, 1985; Sharkey and Folkins, 1985; Smith and McLean-Muse, 1986; Zajac and Hackett, 2002) and even into adolescence (Walsh and Smith, 2002). Similar age-related effects have been reported in studies of nonspeech motor tasks such as reaching, tapping, and pinching (Berthier, 1996; Deutsch and Newell, 2001, 2004; Tingley and Allen, 1975; Yan *et al.*, 2000), suggesting that the phenomenon largely reflects motoric (rather than linguistic) development. This paper will mostly treat speech production variability from a motor control perspective, but results from work such as this may have implications for theories of developmental phonology, and some results from more linguistically oriented work will be addressed briefly below.

Several factors must be considered when evaluating the nature of developmental variability. One issue is that children's motor acts are not only variable on average relative to adults' but also tend to display lower velocities and longer segment or movement durations (e.g., see Chermak and Schneiderman (1986), Eguchi and Hirsh (1969), Kent and Forner (1980), Smith and Goffman (1998), Smith (1978), and Yan *et al.* (2000)]. Since variability, all else being equal, increases with the absolute magnitude of a measure (Ohala, 1975), developmental studies often employ relative measures such as the coefficient of variation, where the standard deviation (SD) is divided by the mean. In a series of studies, Smith and co-workers demonstrated that longer durations account for part but not for all of the higher variabilities seen in temporal measures of children's speech [Smith, 1992; Smith, 1994; Smith *et al.*, 1983; Smith *et al.*, 1996; see also Chermak and Schneiderman (1986)].

Further, age-related differences may vary depending on the articulators and signal types assessed. For example, Green *et al.* (2002) reported that 1- and 2-year-old children have more adultlike variability for the jaw than for the lips. Nittrouer (1993) similarly proposed that young children have more mature control over the jaw than the tongue. Stathopoulos (1995) emphasized that high variability observed at one level of analysis does not necessarily imply high variability on others, even for measures typically thought to be closely related, such as subglottal pressure and speaking sound pressure level. Goffman *et al.* (2002) presented another example of differences across signal types. In a longitudinal case study on speech production after cochlear implantation, these authors found that durational variability in the acoustic signal declined steadily following implantation, whereas kinematic variability first increased before decreasing to more age-appropriate levels. Although these changes over time presumably resulted from the child's changing auditory feedback rather than simple speech motor development, they do indicate that kinematic and acoustic recordings of the same speech events may show different trends.

Smith and Goffman (1998) considered whether temporal and spatial aspects of speech mature at the same rate. Their study used the spatiotemporal index (STI) to quantify lower-lip displacement variability in 4 year olds, 7 year olds, and adults producing repetitions of “buy Bobby a puppy.” As expected, STI results were significantly higher in children than in adults. Since the STI combines the temporal and amplitude aspects of variability into a single number, two subsequent analyses were performed on the time-normalized data to investigate possible differences between these two dimensions. An assessment of the relative timing of velocity peaks in the lowering movements for /a/, /iə/, and /ʌ/ in “Bobby a pup-” revealed a tendency, significant for the 7 year olds but not for the 4 year olds, for children to produce their first and second velocity peaks relatively later in the utterance than adults. In contrast, success rates for a cross-correlation pattern-matching classification on the lower-lip displacement in the syllables “Bob” and “pup” showed no significant age effects. Smith and Goffman (1998) concluded that relative timing in an utterance matures later than spatial displacement in individual syllables. The authors characterized this finding as unexpected since the long time course of physical growth might lead one to expect spatial features to show more protracted development than temporal features. It is interesting to note in this context, however, that a longer acquisitional time frame for temporal aspects of speech production is reminiscent of a theme that has recurred in the developmental phonology literature. Specifically, some child language researchers (Ferguson and Farwell, 1975; Vihman, 1993; Waterson, 1971) have observed that young children’s word productions may contain many articulatory gestures appropriate for the adult model but lack consistent temporal phasing between them.

The methods of the present study were designed to address many of the issues identified in the preceding paragraphs. Specifically, variability measures in FDA were derived relative to the mean and SD for each speaker and consonant; variability was compared across different articulators; and measures were obtained for both temporal and spatial aspects of the data.

C. Interpreting variability

The trends of decreasing motoric variability, higher movement velocities, and shorter durations with age have often been attributed to neurological changes including cortical and cerebellar development, stabilization of electroencephalography (EEG) patterns, and myelination of the motor tracts [e.g., see Kent (1976), Lecours (1975), Tingley and Allen (1975), and Yan *et al.* (2000)]. From this perspective, variability in children may result from increased sensorimotor noise in a less stable system. Contemporary models of motor control and sensorimotor integration often include explicit noise terms [e.g., see Körding and Wolpert (2004), Todorov and Jordan (2002), and Vetter and Wolpert (2000)] and Berthier (1996) successfully modeled infant reaching data using a high level of random noise.

Other work suggests that random neurological noise may not differ as a function of age, however. First, extensive

cross-subject differences have been observed in developmental studies, with some children having rather adultlike values for some variability measures (Smith, 1995; Smith *et al.*, 1996; Smith and McLean-Muse, 1986; Stathopoulos, 1995). Such individual variation would seem to argue against a simple additive noise factor that differs between children and adults (Stathopoulos, 1995). Further, Deutsch and Newell (2001, 2003, 2004) found that frequency analyses of finger pinching did not show age differences in random noise; rather, children’s movements had reduced dimensionality and more energy in lower frequencies. These authors proposed that children differ from adults in being less skilled at meeting task demands and integrating sensory feedback during ongoing movement control.

The motor learning literature indicates that practicing a task leads to faster, more accurate, and less variable performance (Hatze, 1986; Liu *et al.*, 2006; Schulz *et al.*, 2000; Sossnik *et al.*, 2004). Thus, similar developmental patterns for speech durations and variability (Eguchi and Hirsh, 1969; Kent, 1976) may simply result from children being less practiced speakers. Deutsch and Newell (2004) observed that, with practice, some children approached adultlike values of force variability in a nonspeech (pinching) task and, moreover, that practice effects accounted for more variance than chronological age. Age effects can also be reduced by manipulating task demands, including accuracy requirements, force levels, and presence of feedback (Deutsch and Newell, 2001, 2003, 2004; Yan *et al.*, 2000).

Finally, both speech and nonspeech studies indicate that production variability may increase during periods of learning or skill acquisition when the motor system is in transition to a new mode of organization (Berthier, 1996; Goffman *et al.*, 2002; Liu *et al.*, 2006; Thelen *et al.*, 1996; von Hofsten, 1989). In this view, specific patterns of variability arise as individuals seek successful motoric strategies for a given goal (Jensen *et al.*, 1995). Goffman *et al.* (2002) [see also Smith and Goffman (1998)] noted that greater plasticity in the developing nervous system may facilitate task-space exploration and be a sign of a flexible and adaptive organization. This implies that the underlying causes of variability may have functional utility during development.

Although variability may ultimately derive from many sources—and this can lead to some indeterminacy [see Hatze (1986)]—sophisticated experimental paradigms may yield greater insight into the likely sources of developmental variability for particular tasks, as well as the nature and time course of age-related changes. As outlined in the next section, FDA allows token-to-token variability to be partitioned into spatial and temporal domains. This, combined with a comparison across articulators, may speak to the general debate about the extent to which developmental variability reflects a general random noise factor versus individualistic task-specific patterns.

D. FDA as a technique for exploring variability

FDA was developed by Ramsay and Silverman (1997) as a means of analyzing time-varying signals. A distinctive feature of FDA is nonlinear time warping to bring multiple

signals into closer alignment with an average signal. These methods may be used in service of many questions. For example, previous researchers have used FDA to explore the underlying dimensionality of speech movements (Ramsay *et al.*, 1996), how speech timing changes near phrasal boundaries (Lee *et al.*, 2006), the most appropriate method for calculating the harmonics-to-noise ratio of voice signals (Lucero, 1999), and the degree to which kinematic variability in consonants depends on the articulators involved in forming the constriction (Lucero and Löfqvist, 2005). The present work capitalizes on the fact that FDA allows variability to be decomposed into two dimensions: The magnitude of the time warping functions used to align the data provides an index of temporal (phasing) variability, whereas the remaining variance in the aligned waveforms represents amplitude variability once most phasing variability has been factored out. One question of this work was to determine whether these two forms of variability show parallel developmental trends.

E. Current study

In summary, the purpose of this study was to investigate speech production variability in children using two strategies. First, variability was explored across oral, laryngeal, and combined oral-laryngeal actions. It may be that stable control is acquired earlier over one of these articulatory systems than the other. Alternatively, speakers may differ in which system shows more stable behavior at a particular time. Comparing the variability patterns for the three consonants was intended to elucidate the ways in which children, both individually and as a group, learn to control lingual and laryngeal articulations alone and in combination. Variability in the three consonants was assessed quantitatively using FDA methods and qualitatively, over time, in a vowel-consonant-vowel (VCV) sequence. Second, decomposing variability into distinct phasing and amplitude components permitted exploration of whether these two dimensions have similar developmental profiles. *A priori*, it seemed sensible to predict that both phasing and amplitude variabilities would be higher in children than in adults, but whether the two measures have equivalent developmental time courses was an open question. Differences in variability patterns for the three consonants or for temporal and amplitude variabilities would suggest that a simple random noise factor is insufficient to explain variability differences between children and adults. Thus, the specific questions of this study were as follows: (a) How does variability compare across /h s z/ within and across age groups? (b) Do amplitude and phasing variabilities develop in parallel for this set of consonants?

II. METHODS

A. Speakers

Ten children in each of two age groups and ten women were recorded. The younger children ranged in age from 4.5 to 6.0 years ($\mu=5.07$, henceforth called 5 year olds). The older children were 8.9–10.8 years of age ($\mu=9.86$, henceforth called 10 year olds). The child age groups were chosen so as to obtain data from children (a) as young as

possible, given the demands of the experimental task, and (b) as old as possible without introducing gender differences as a function of puberty or minimizing the likelihood of observing age effects. Some measures of variability in children may approach adultlike levels at around 12 years of age (Eguchi and Hirsh, 1969; Lee *et al.*, 1999). Each of the child groups had five boys and five girls.

All participants were native speakers of American English from the New York City metropolitan region. The children had normal developmental histories as ascertained by parental report, including no history of intervention for speech or language problems. Parents of the 10 year olds also verified that their children had not yet begun to show pubertal changes. A short narrative or conversational sample was obtained from each child speaker (targeting 50 or more utterances; all included at least 35) and was analyzed to confirm age-appropriate language skills, as measured by mean length of utterance (MLU) and syntactic complexity (Hughes *et al.*, 1997; Miller and Chapman, 1981). Finally, all children passed a hearing screening to establish binaural thresholds at 25 dB or less over 250–8000 Hz.

The adult female group consisted of mothers of the child participants. All had normal speech, language, and hearing histories by self-report and speech characteristics within normal limits as judged informally by the first author. Mothers were used as the comparison group for two reasons. First, some of the younger children were more comfortable doing the experiment after watching their mothers perform the task. Second, this minimized dialectal variation, an important consideration with relatively small group sizes. The adult group was further restricted to women because they have laryngeal and vocal-tract proportions more similar to children's than do men. Although restricting the adult group to women may limit the generalizability of the adult data somewhat, no evidence exists to indicate that men and women differ in their variability characteristics for repeated productions of speech sequences, so this limitation should not affect general conclusions regarding age effects.

The experimental protocol was approved by an Institutional Review Board and followed standard procedures for the protection of human research participants. Before recording, parents provided informed consent for their own and/or their child's participation, and all children provided assent. All participants were naive as to the purposes of the experiment.

B. Speech materials

Speakers were recorded producing approximately 25–30 repetitions of the utterances “Poppa Hopper,” “Poppa Sapper,” and “Poppa Zapper.” The targets /h s z/ thus had a common preceding context (bilabial closure followed by schwa) and a low vowel following (/a/ or /æ/). The utterances were elicited by means of colored drawings of a bunny carrying out activities related to his name (*viz.*, hopping, draining sap from a tree, and getting zapped by a lightning bolt). The target consonant was also prominently displayed on the bunny's shirt, and his name was printed at the top or bottom of the picture. A research assistant introduced the

pictures to speakers before recording, pointing out the salient features of the picture. For the younger (preliterate) child group, picture naming was elicited immediately following the description (e.g., “This guy is getting SAP from a tree, so he’s Poppa SAPPER. Can you say that?”). Speakers were told that when they saw a picture, they should say the bunny’s name and then repeat it five times.

During recording, the research assistant presented pictures for naming and counted off on her fingers as the speaker repeated the utterance four to five times. Since the utterance list was very short, the target consonant was given on the bunny’s shirt, and the bunny’s name was written on the picture, all adults and most child speakers were usually successful in naming the picture accurately on the first presentation. For some of the 5 year olds, additional prompts were given if the child’s first production was not perceptually adequate (e.g., “He’s hopping. What’s his name?”). Pictures were presented five times per recording session, in randomized order. For analysis, tokens were excluded if there was a pause before the target consonant or if the target consonant was not perceptually accurate in place, manner, and voicing. No speaker was included in the final data set if s/he did not produce at least ten analyzable tokens of each consonant. The average number of tokens analyzed for each consonant and speaker was 27 (SD=5).²

Most often the speakers produced all repetitions on a single breath, but a few of the 5 year olds occasionally inspired between repeated productions. Speech rate was not explicitly controlled, but any speaker who at first used a slow careful speaking style was encouraged to speak more naturally and casually so that there were no perceptually clear pauses between words within a target utterance.

C. Equipment and recording

Oral airflow signals were collected using an undivided Glottal Enterprises pneumotachograph appropriately sized for the speaker. The airflow data were digitized at 10 kHz onto a PowerLab recording system connected to a Macintosh laptop computer. An audio signal was simultaneously recorded to the PowerLab at a sampling rate of 20 kHz using an AKG 420 microphone hung around the speaker’s neck. The microphone signal was used to determine that the target consonant was accurate at the level of broad transcription and that the speaker did not pause before the target consonant. All analysis for this work was performed on the airflow data. Flow signals were calibrated before or after each recording session using a rotameter.

D. Preliminary processing

Initial data processing used smoothing and differential routines in the Chart™ software that accompanies the PowerLab. Since the articulatory activities of interest, namely, glottal abduction and supraglottal constrictions, are relatively low in frequency, the airflow data were smoothed twice using a triangular window of 201 points (approximately 20 ms, or a 50 Hz low-pass cutoff). This removed the rapid oscillations associated with vocal fold vibration. First and second derivative (velocity and acceleration) signals were obtained from

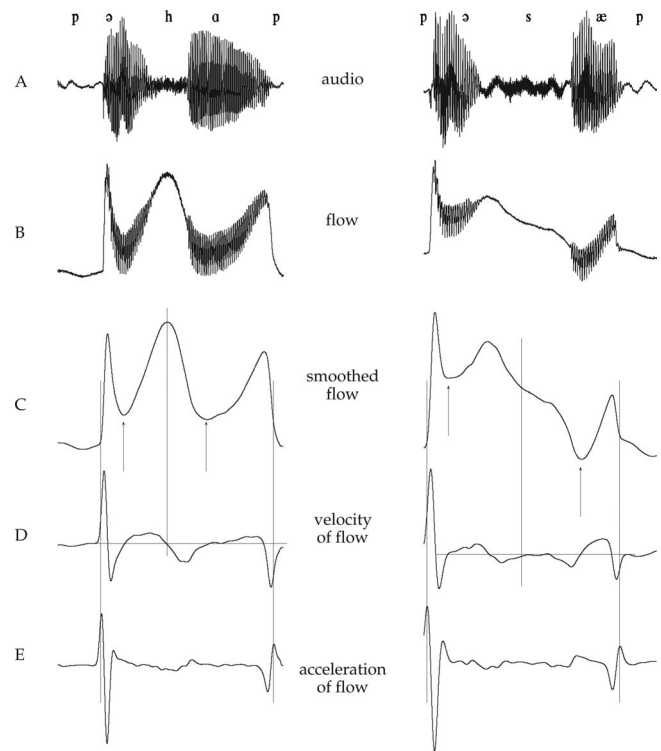


FIG. 1. Sample tokens of /pəhap/ and /pəsæp/ from a 5-year-old boy. The time axis is 500 ms for both sequences. (A) Audio signal. (B) Original airflow signal. (C) Smoothed airflow signal. The arrows show the region over which the final amplitude index was calculated (viz., between the flow minima of the two vowels). (D) Velocity of smoothed airflow. Zero-crossings in this signal were used to identify flow peaks in /h/ and minima in /s z/. (E) Acceleration of smoothed airflow. The peaks in this signal reflecting release and closure of the preceding and following /p/s marked the boundaries of the FDA warping window.

the smoothed flow signals and smoothed iteratively with 201-point windows (same specifications as above) so that peaks and zero-crossings could be obtained easily. Typically two passes of filtering were sufficient for this purpose, but for a few speakers whose signals were noisier, a third round of filtering was implemented to remove remaining high-frequency oscillations that complicated the semiautomatic event selection. Individual tokens of /h s z/ were then labeled as follows (see Fig. 1): The peak of the smoothed flow signal in /h/, representing the maximum abduction, was identified from zero-crossings in the velocity signal. Flow minima during the oral constrictions for /s z/ were likewise identified using the velocity zero-crossing. Finally, the second-derivative peaks representing the flanking /p/ release and closure were identified. All peak and zero-crossing labeling was performed with reference to the original flow signals to ensure that appropriate locations were chosen. To prepare the data for FDA processing, individual tokens were extracted by selecting a window centered around the /h s z/ peak or valley, long enough so that it included both /p/ acceleration peaks. These signals (smoothed flow data) were saved as text files and imported into MATLAB. In MATLAB, the data were trimmed so that each token began and ended with consistent aerodynamic events, namely, the second-derivative peaks reflecting /p/ release and closure. This defined the VCV sequence used in subsequent processing.

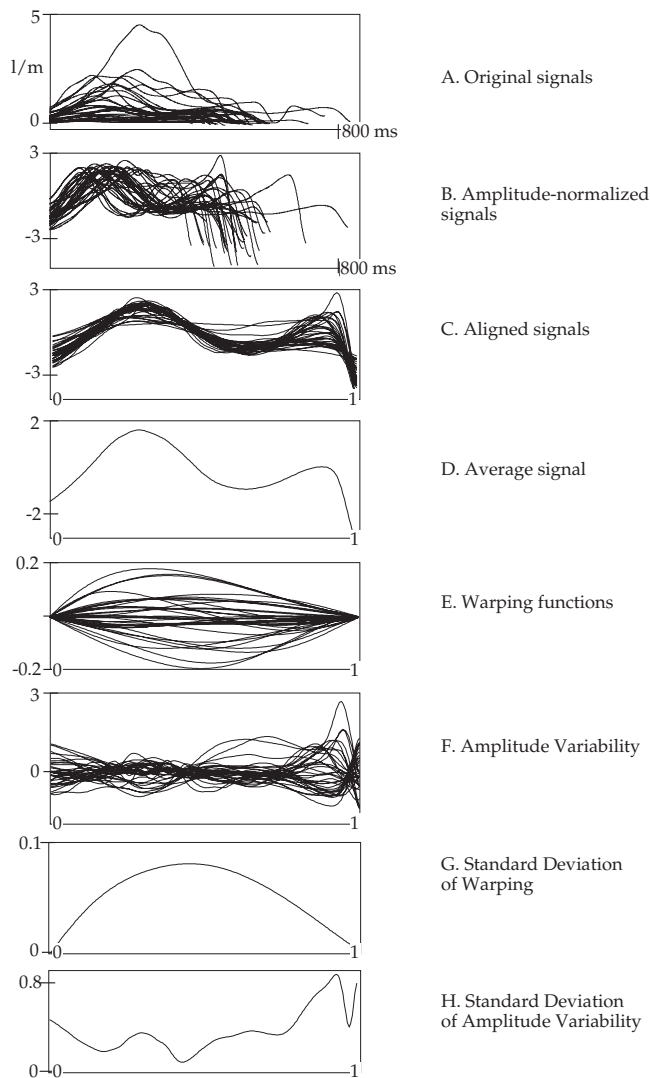


FIG. 2. An illustration of the processing steps for all tokens of /h/ in speaker 5m1 (a 5-year-old boy). Panels top to bottom: (A) All original productions of /h/ not normalized for time or amplitude. (B) Amplitude-normalized tokens. After amplitude normalization, the y-axis is unitless. (C) Time-normalized (i.e., aligned) tokens. After time normalization, the x-axis is unitless. (D) Average of all tokens after time and amplitude normalization. (E) Warping functions for all tokens. (F) Amplitude variability (SDs) for all original records relative to average. (G) SD of warping functions over normalized time. (H) SD of amplitude variability functions over normalized time.

E. FDA processing

FDA methods may vary somewhat depending on the goals of the work, but the general purposes are to reduce the data (so as to facilitate processing) and, most critically, to align the signals in time using nonlinear time warping. The magnitude of the time warping required to bring the set of signals into alignment (as determined by a cost-minimization function) represents phasing variability, and the variation around the average obtained from the aligned data provides a measure of amplitude variability, which is relatively independent of temporal variability. The specific processing steps used in the current work, for all tokens of a consonant within a speaker, were as follows (see Fig. 2):

(a) The original tokens [panel (A) in Fig. 2] were vertically

aligned by subtracting the mean of each token and amplitude normalized by dividing each token by its SD [Fig. 2(B)]. This normalization corrects for greater variability resulting from increased magnitudes alone (see Ohala, 1975).

(b) The amplitude normalized tokens were processed using the FDA package of MATLAB functions given by Ramsay and Silverman at <ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns> [see also Ramsay and Silverman (1997, 2002)].

First, the tokens were put into functional form by expanding them into a basis of 42 fourth-order B-splines. Using an order of 4 ensures that the signals will produce smooth acceleration functions. At the same time, their duration was set to a normalized time span of [0, 1]. This number of basis functions was sufficient to capture details of the tokens' waveshapes, while applying a light smoothing, as assessed by a visual inspection of the results.³

The tokens were then aligned in time by nonlinear time warping. In this process, nonlinear transformations of time $h(t)$ (warping functions) are computed for each token so as to minimize the following measure of shape similarity:

$$\text{MINEIG}_\lambda(h) = \text{MINEIG}(h) + \lambda \int w^2(t) dt, \quad (1)$$

where w is the relative curvature of the warping function, λ is a roughness penalty coefficient imposed on the warping function, and $\text{MINEIG}(h)$ is the logarithm of the smallest eigenvalue of matrix

$$\begin{bmatrix} \int [x_0(t)]^2 dt & \int x_0(t)x[h(t)]dt \\ \int x_0(t)x[h(t)]dt & \int x[h(t)]^2 dt \end{bmatrix}, \quad (2)$$

where $x(t)$ is the token being normalized and $x_0(t)$ is the target for the time warping. For these analyses, λ was set to 10^{-5} , ensuring smooth warping functions of time. Further, the warping functions were expressed in a basis of ten fourth-order B-splines, following previous work (Lucero *et al.*, 1997). The computation of the optimal warping functions was performed in two steps. A first approximation was obtained using the mean of the tokens as the target [$x_0(t)$ in Eq. (2)]. A new target was then obtained as the mean of the aligned tokens (computed in the first step), and the warping functions were refined in a new run of the algorithm. Panel (C) in Fig. 2 shows the data after applying the time warping, and panel (D) shows the average calculated from the aligned data.

(c) To assess the degree of temporal variability of the signals, the SD of the time warping functions over time was obtained [see panels (E) and (G)].

(d) To assess the amount of amplitude variability in the aligned signals, the waveform noise (deviation from the average signal) was obtained for each token [panel (F)], and the SD of this set of functions was obtained [panel (H)].

Early analyses indicated that speakers (women as well as children) tended to have high variability at the beginning and/or end of their records [note the late peak in the amplitude SD function in Fig. 2 (H)], resulting from airflow variation related to the preceding and following /p/. Since the goal of this analysis was to characterize the VCV sequence and not the aspects of /p/ *per se*, the final amplitude index was calculated over the region defined by flow minima in the vowels flanking the target consonant. This region is indicated by arrows in Fig. 1(C).

The final analyses were based on the following data for each consonant and speaker: (a) Average smoothed airflow signals for /h s z/. (b) Two SD functions over time (amplitude and time warping). (c) The mean of the time warping SD [i.e., the mean of the signal in Fig. 2(G)] obtained over the entire VCV sequence (henceforth called the time warping index). (d) The mean of the SD of the waveform noise [Fig. 2(H)] calculated over the region indicated in Fig. 1 (henceforth called the amplitude index).

The time warping index provides a measure of temporal variability in the airflow peaks and valleys across tokens. The amplitude index represents the degree to which the magnitude of oral flow varied across tokens after factoring out temporal variability. Flow rates, in turn, reflect laryngeal and oral events: Laryngeal abduction will increase airflow rates, all else being equal, whereas supraglottal constrictions will reduce airflow. Thus, the two indices measure different types of production variability: gestural phasing versus magnitude of abduction or tongue-tip constriction. These summary measures permit statistical testing of age and consonant effects on variability. The amplitude and time warping functions over time were also assessed qualitatively to gain insight into individual patterns of variability.

F. Statistics

The amplitude and time warping indices for the 3 consonants and 30 speakers were entered into two-way analyses of variance (ANOVAs) using STATVIEW, with independent variables of age group and consonant. Since two separate ANOVAs were being run, the α -level was set to 0.025. In cases of significant main effects or interactions, *post hoc* Fischer's PLSD tests were run. Correlations (Pearson's r) were also carried out to evaluate the relationship between amplitude and time warping indices within speakers.

III. RESULTS

A. Age and consonant effects on amplitude and time warping indices

As indicated above, the amplitude and time warping indices are useful for performing statistical tests of how variability differs as a function of age, across consonants, and in temporal and amplitude domains. Figure 3 presents the amplitude (top) and time warping (bottom) values for all speakers. The average amplitude and warping indices are given for the three age groups in Fig. 4. Tables I and II summarize the results of the ANOVAs and the *post hoc* tests, respectively.

As expected, both indices decreased as a function of age. The main effects of age and all pairwise comparisons were

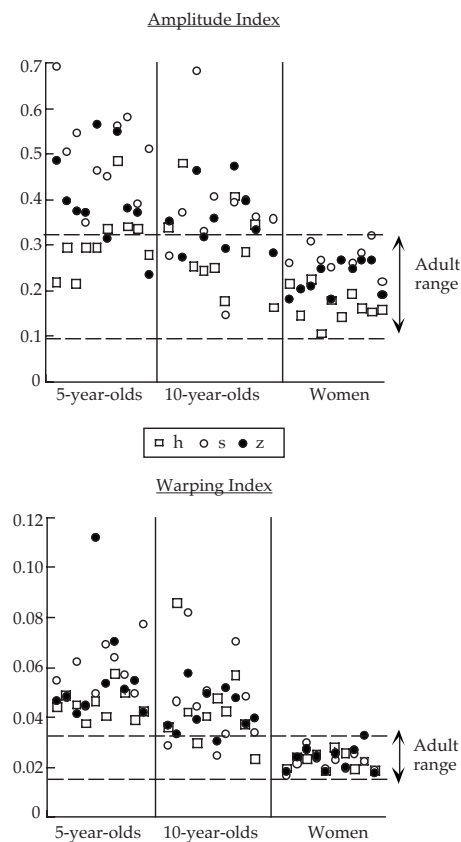


FIG. 3. Amplitude (top) and warping (bottom) indices for all speakers. Within the child groups, the leftmost five speakers are the girls; the rightmost are the boys. An open square represents /h/; an open circle represents /s/; a filled circle represents /z/.

significant for both measures (see Tables I and II). As shown in Fig. 3, however, the two indices have somewhat different developmental time courses. For the amplitude index, several values for both 5 and 10 year olds fall within the adult range (8 and 12 values out of 30, respectively). For the time warping index, relatively few of the 10 year olds' values (5 of 30) and *none* of the 5 year olds' values fell within the adult range. Thus, for this speech context, gestural magnitude appears, on the whole, to reach adultlike stability levels earlier than gestural phasing.

The main effect of consonant was significant for the amplitude index but not for the time warping index (see Table I). *Post hoc* tests showed that /h/ had lower amplitude variability than both /s/ and /z/ (see also Fig. 4). Pairwise comparisons for /s/ and /z/ were also significant. Figure 4 suggests that, qualitatively, the differences among the three consonants were the most extreme for the 5 year olds, but the age-by-consonant interaction was not significant presumably because of high within-group variability among the children combined with a modest sample size. Despite the mean differences between the three consonants (top panel of Fig. 4), it is evident from Fig. 3 (top panel) that not all speakers demonstrate the pattern of $h < z < s$ in their amplitude index. Individual patterns of variability are considered in Sec. III B.

To determine the relationships between the two indices, correlations were run for the whole data set (Fig. 5, top

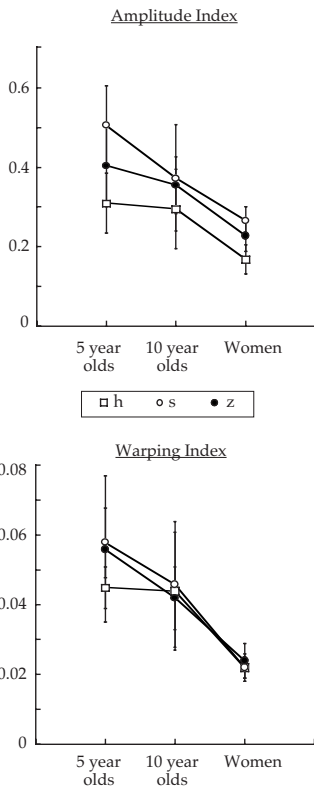


FIG. 4. Group averages for amplitude and warping indices for the three consonants. An open square represents /h/; an open circle represents /s/; a filled circle represents /z/.

panel) and then divided by consonant (middle) and age (bottom). For the combined data (top panel), the correlation was significant ($r=0.791$, $p<.0001$), with $r^2=0.625$. Thus, the two measures were moderately predictive of one another overall, but not totally redundant, since nearly 40% of the variance was unaccounted for. The middle panel of the figure shows that the correlation held about equally across the three consonants: All r -values were significant ($p<0.0001$), in the range of 0.76–0.86, and the slopes (m -values) were similar across /h s z/. When the data were split by age, however, clear group differences emerged (bottom panel). In particular, the two indices did not correlate for the women ($r=0.23$, $p=0.2235$), whereas they did for both groups of children. In adults, then, temporal and amplitude variabilities seem to be entirely independent for these speech sequences.

TABLE I. Results of ANOVAs.

	DFs	F	p
Amplitude index			
Age	2	38.372	<0.0001
Cons	2	16.495	<0.0001
Age × Cons	4	1.486	0.2140
Residual	81		
Warping index			
Age	2	49.014	<0.0001
Cons	2	1.188	0.3100
Age × Cons	4	1.155	0.3369
Residual	81		

TABLE II. Results of *post hoc* tests.

	Amplitude index <i>p</i> -value	Warping index <i>p</i> -value
Age effect		
Adults vs. 5 year olds	<0.0001	<0.0001
Adults vs. 10 year olds	<0.0001	<0.0001
5 year olds vs. 10 year olds	0.0031	0.0064
Consonant effect		
/h/ vs. /s/	<0.0001	0.1426
/h/ vs. /z/	0.0014	0.2692
/s/ vs. /z/	0.0183	0.7140

B. Individual patterns of consonant-related variability

This section explores within-speaker patterns in amplitude and warping over time. This qualitative analysis supple-

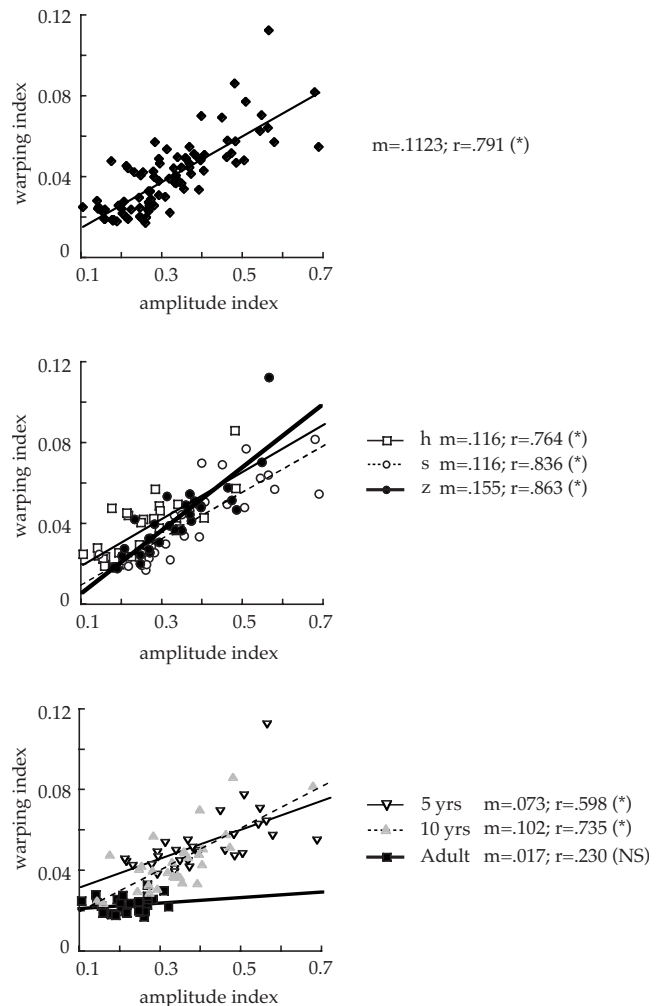


FIG. 5. Correlations between warping and amplitude indices for all data (top), the data split by consonant (middle), and the data split by speaker group (bottom). When split by consonant, open square, solid line represents /h/; open circle, dotted line represents /s/; and filled circle, heavy solid line represents /z/. When split by age group, open triangle, solid line represents 5 year olds; gray triangle, dotted line represents 10 year olds; and filled square, heavy line represents adults. Asterisks after the r -values indicate that the correlation was significant at $p<0.01$; NS indicates that the correlation was not significant. As noted for Fig. 2, the scales are unitless.

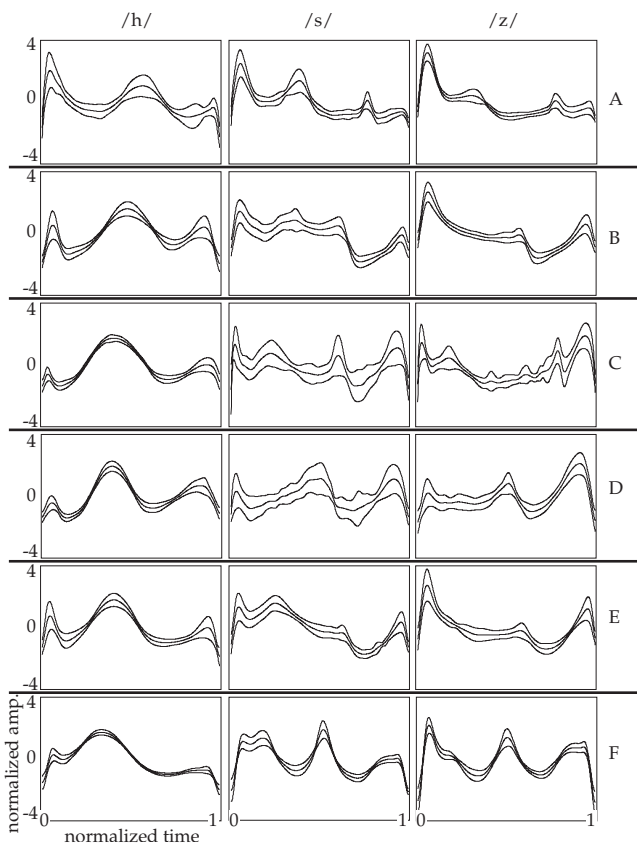


FIG. 6. Speaker differences in amplitude variability over time in /h/ (left), /s/ (middle), and /z/ (right). Top to bottom: row A, speaker 10F2 (10-year-old female). Row B, 5M1 (5-year-old male). Row C, 5F1 (5-year-old female). Row D, 10F3 (10-year-old female). Row E, 5M4 (5-year-old male). Row F, AF7 (adult female). Rows A and B show speakers who had relatively high variability in /h/ and /s/ but less in /z/; C and D represent speakers with relatively high variability in /s/ and /z/ but less in /h/; E and F show speakers where variability is similar across the three consonants.

ments the quantitative index analysis above by providing a dynamic picture of amplitude and phasing variability over the VCV sequences. A key question is how variability over time compares among the three consonants.

Time-varying averages ± 1 SD for a subset of speakers are given in Fig. 6. Recall that these average and SD functions were calculated on the aligned data. Thus, amplitude variability around the /h/ flow peak relates to differences in abduction degree after variations in temporal phasing have been factored out. Variability around the low-flow region of /z/ arises from differences in tongue-tip constriction degree. The typical profile for /s/ is an overall increase in airflow, resulting from vocal fold abduction, with a superimposed valley, representing the tongue-tip constriction (Klatt *et al.*, 1968). Of particular interest here is how variability during /s/ relates to that in /h/ and /z/. If both /s/ and /z/ show higher variability than /h/, it suggests that the speaker has less stable control over lingual constrictions than over laryngeal abduction. On the other hand, high variability for /s/ and /h/ but not for /z/ would imply more variable laryngeal control than supralaryngeal. Finally, it could be that variability in /s/ reflects supraglottal and glottal control to similar degrees. Figure 6 indicates that all three of these options may be observed across speakers, adults as well as children. For the two

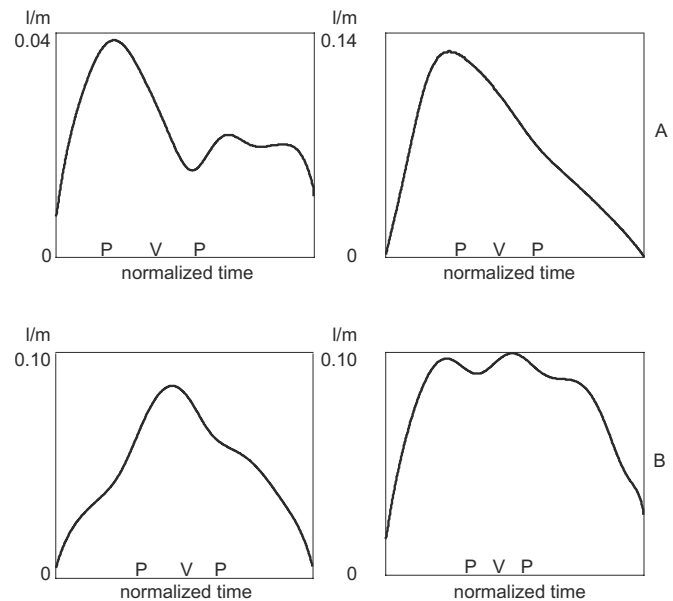


FIG. 7. Speaker differences in the location of maximum warping in /s/. The P V P notations in each figure show the temporal location of the two peaks and the valley characteristic of airflow during /s/. (A) Left: speaker AF7 (adult female); right: 10M3 (10-year-old male). (B) Left: speaker 5F6 (5-year-old female); right: 5M5 (5-year-old male). In all graphs, the x -axis is normalized time over the range of 0–1.

speakers shown in rows A and B, variability is high during /h/ (left panel) and /s/ (middle panel), whereas /z/ production (right panel) is more consistent. For the speakers in rows C and D, /s/ and /z/ have similar variability, whereas /h/ is less variable. Finally, the two speakers in rows E and F have similar variability across the three consonants; that is, /s/ appears to have about equal variability in laryngeal and supralaryngeal aspects of articulation.

Speakers also differed in terms of where their signals showed the most extreme warping. In the FDA warping algorithm, the edges of the records (that is, normalized times 0 and 1) are fixed in time, so phase variability is zero at those times. Between the beginnings and ends of the signals, however, considerable variation was observed, with no clear patterns based on age or consonant. Warping functions for a subset of speakers producing /s/ are given in Fig. 7. (Similar degrees of cross-subject variability were observed for the other two consonants.) The notations at the bottom of each graph indicate the relative temporal locations of the two peaks and the valley characteristic of /s/ production. For some speakers (row A), the most extreme warping was required early in the VCV during the increase in airflow that reflects abduction for /s/. In these data, it happened that no speakers had a clear “late” peak for /s/, but such cases were observed in /z/ and /h/. Most typical was for the warping peak to occur near the middle of the signal, here, around the time of maximum tongue-tip constriction for /s/, but speakers differed in whether there was a fairly short-term peak versus a more extended plateau (row B). In short, the dynamic patterns of amplitude and phasing variabilities revealed considerable cross-speaker differences.

IV. DISCUSSION

A. Temporal and spatial variability

1. Similarities and differences between amplitude and time warping measures

Past research on speech and other motoric behaviors has found higher token-to-token variability, on average, in children compared to adults. Many authors have appealed to aspects of neurological development to argue that children have more random sensorimotor noise than adults (Kent 1976, Lecours 1975, Netsell 1981, Tingley and Allen 1985, and Yan *et al.* 2000). At the same time, extensive differences have been observed within groups for specific variability measures and within individuals across measures (Smith, 1995, Smith and McLean-Muse, 1986, Stathopoulos, 1995, and Tingley and Allen, 1975), suggesting that variability in children may not be adequately characterized as simple random noise.

The use of FDA here allowed the production variability to be partitioned into two components. It was hypothesized that both amplitude and warping variabilities would be higher in children than in adults, but it was unclear whether the two indices would change in parallel over age. Significant group effects emerged for both measures. Thus, consistent with previous studies, children, as a group, were more variable than adults. Yet the indices also differed in two ways: (a) whereas there was considerable overlap between the 5 year olds' and women's amplitude indices, there was none at all for the warping index, and (b) a significant consonant effect was observed for the amplitude index but not for the warping index. The specifics of the amplitude and time warping index results are considered in more detail in Secs. IV A 2 and IV A 3. Here, the observation of primary interest is that the two measures do not show parallel developmental time courses.

Smith and Goffman (1998) also suggested that the temporal and spatial aspects of speech production might mature at different rates. As described in Sec. I B, their study measured lower-lip displacement over a sentence, carried out linear normalization, and then assessed (a) the relative timing of velocity peaks for three vocalic sequences and (b) the automatic classification success for waveform displacement in two consonant-vowel-consonant (CVC) syllables. Significant age differences were found for the relative timing measures but not for the classification success. There are many methodological differences between this work and that of Smith and Goffman (1998), including normalization techniques, ways of assessing temporal versus spatial variability, the duration of the speech sequences, and the data type (kinematics versus aerodynamics). Despite these differences, both studies found that temporal control in speech production, at least over sequences of a disyllable or longer, matures over a longer time period than amplitude variability or gestural shape.

These differences in temporal and spatial aspects of speech support the use of more sophisticated analyses of variability, such as those provided by FDA, to elucidate the nature of developmental changes across multiple dimensions. The finding that the two indices were not correlated in

women further justifies considering them separately. Significant correlations in the children but not in the women could indicate greater overall noise in developing speakers for both temporal and amplitude domains. Yet such noise would not be truly random insofar as it is correlated across gestural magnitude and gestural phasing. While the current findings do not rule out a higher random noise factor in children, they do suggest that such a factor, alone, cannot explain all developmental differences.

As mentioned in Sec. I B, some developmental phonologists have also suggested that children may acquire articulatory gestures before they learn the appropriate phasing among them (Ferguson and Farwell, 1975; Vihman, 1993; Waterson, 1971). It must be noted that these observations were made on children much younger than those here and used a very different level of analysis, namely, phonetic transcription. It is not clear, therefore, whether the current results truly represent an example of the same phenomenon. More instrumental data, including studies of younger children, with varying speech sequence lengths and a wider variety of phonetic contrasts, are needed to ascertain whether the present findings are specific to these speech materials and methods or whether they indicate something more general about the nature of phonetic learning during childhood.

It is also unclear to what extent the current findings reflect purely motoric aspects of speech production versus ongoing linguistic development. Although children of the age studied here are usually considered to have a fairly complete sound inventory for their language, the details of phonological representations may well be refined over a longer time period. Further, speech differs from many tasks used in studies of nonspeech motor control in being a highly practiced skill. Again, assessing the generality of these results will require future studies that use similar analysis methods to explore a wider range of tasks, sampling across multiple articulators. Comparisons between speech perception and production abilities in the same children may also help to differentiate linguistic and motoric domains (see, e.g., Nittrouer and Studdert-Kennedy, 1987; Nittrouer *et al.*, 1989).

2. Amplitude index

The amplitude index in this study represents variability in airflow once phasing variability has been largely removed through nonlinear warping. The consonant effect for this measure indicates that airflow management is more consistent in /VhV/ sequences than in /VzV/ or /VsV/ for both adults and children. Since variability may increase with task demands (e.g., Yan *et al.*, 2000), these data provide some evidence that the lingual fricatives impose a greater motoric burden than the glottal fricative. Further, variability was highest, on average, for /s/, suggesting that the combination of abduction and supraglottal constriction, the most articulatorily complex sequence of the three, is also the most variable in gestural magnitudes. Although the warping index did not show significant consonant effects, an inspection of Fig. 4 (bottom panel) does suggest a possible qualitative difference in the 5 year olds, with less phasing variability in /h/

than in /s/ and /z/. More data might reveal that young children do not achieve equivalent phasing control of all articulators at the same time.

3. Time warping index

Many past speech studies have observed that individual children may be rather adultlike on some variability measures (Smith, 1995; Smith and McLean-Muse, 1986; Smith *et al.*, 1996; Stathopoulos, 1995). The 5 year olds' time warping data obtained here appear to represent an exception to that general conclusion: None of their 30 index values fell into the adult range.

One possibility that should be ruled out is that this finding results from some kind of processing artifact. Longer segment and syllable durations in children's speech compared to adults' (e.g., see Eguchi and Hirsh, 1969; Kent and Forner, 1980; Smith, 1978) are a potential source of artifact. Specifically, greater time warping for children's data could reflect a greater error "buildup" over a longer time sequence. Two considerations suggest that longer durations in children, all else being equal, should not bias them to higher warping indices.

First, as described in Sec. II E, putting the data into functional form involves expressing them in terms of a fixed number of spline functions. This effectively low-pass filters the data at a frequency that depends on the length of the record: longer records are filtered at lower frequencies. This should actually predispose signals with longer durations to have lower overall shape variability. (In this respect, FDA shares a drawback observed by Lucero *et al.* (1997) for one method of computing the STI.) Second, a short formal test on the effect of record length indicated that this factor, alone, does not yield higher amplitude or time warping indices. One woman's productions of /z/ were selected randomly from the data pool and manipulated as follows: The mean of the records was adopted as an initial pattern, and its length was scaled over a range of 0.5–5.0 of the original duration by linear interpolation. Random phase and amplitude variability components were then introduced into the signals. These components were generated from random time series [see Lucero (2005)] setting their SDs to 0.02 (for phase variability) and 0.2 (for amplitude), similar to the actual measured values for this speaker. Finally, the FDA algorithm was applied to this modified set of data, and warping and amplitude indices were obtained. The results, given in Fig. 8, indicate that simply increasing signal duration does not lead to higher variability indices but indeed, on the whole, to lower ones—especially for the time warping. More work is needed on establishing the precision of FDA computations as a function of signal parameters such as duration, but for present purposes it appears safe to conclude that higher time warping indices in children are not an artifact of longer durations but instead indicate real developmental differences.

B. Individual differences

Despite the general group tendencies revealed by the statistical results, the time-varying plots of mean /h s z/ \pm 1 SD in Fig. 6 indicate considerable variation in cross-

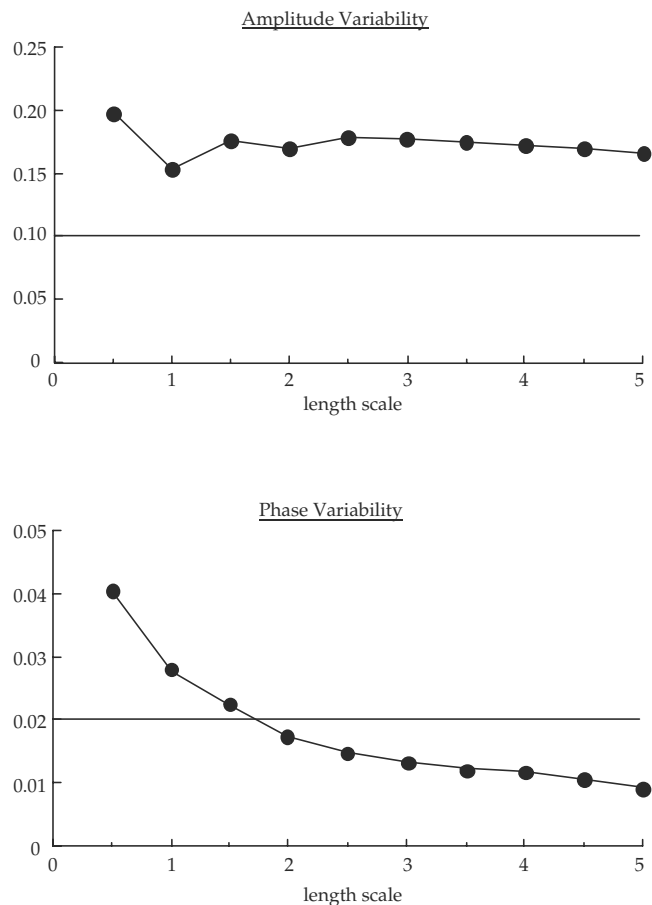


FIG. 8. Amplitude (top) and warping (bottom) indices as a function of increasing signal length.

consonantal patterns across speakers. Some participants had comparable variability in /s/ and /h/ (which have similar laryngeal requirements), whereas others had similar patterns in /s/ and /z/ (which have comparable tongue-tip activity). Further, the peak of the time warping SDs (Fig. 7) occurred in varying locations in the VCV sequence across speakers. Although these are qualitative assessments, they lend strength to the general conclusion that higher variability in children cannot be modeled solely as a higher noise factor that is added to all sound productions across the board. Within-speaker variability differed as a function of the articulatory task (i.e., consonant), and temporal variability was localized in the signal in speaker-specific locations. It appears that children adopt individual strategies for achieving speech goals, as has been observed for adults (Johnson *et al.*, 1993; Koenig *et al.*, 2005; Raphael and Bell-Berti, 1975).

V. CONCLUSIONS

Numerous studies have reported developmental differences in token-to-token variability. A common interpretation has been that random sensorimotor noise is higher in children than in adults, but many authors have drawn on individual differences or theoretical perspectives to emphasize individual learning and exploration. In this work, FDA was employed to decompose variability into two different components and to explore how variability was distributed over the time-normalized VCV sequences. Two findings argue

against attributing children's higher speech production variability to a single random noise factor: Temporal and amplitude variability did not display the same developmental time course, and considerable individual variation was observed in which consonants were the most variable and where in the disyllable the temporal variation was most extreme.

Perceptually adequate productions of /s/ and /z/ occur relatively late for some children. In the current data, /s/ and /z/ were more variable on average than /h/ for all age groups, and /s/ had the most amplitude variability of all, consistent with the prediction that the combination of laryngeal and supralaryngeal articulations for this sound presents an additional level of articulatory complexity. In spite of these group tendencies, some speakers showed different patterns. In particular, a comparison of variability across /h s z/ suggested that laryngeal actions make appreciable contributions to /s/ variability in some speakers.

Little work exists on how children learn to combine laryngeal and oral actions for voiceless fricatives. More data, across a wider range of ages and using a broader variety of speech tasks, are needed to determine the acquisitional time course for control of the larynx as compared to supralaryngeal articulators. Even if general trends emerge, however, the results here suggest that production consistency can differ widely across articulators in speaker-specific ways.

ACKNOWLEDGMENTS

Preliminary analyses and portions of this work were presented at the 140th and 143rd meetings of the Acoustical Society of America, the First Pan-American/Iberian meeting on Acoustics (Cancun), the 7th International Conference on Spoken Language Processing (Denver), and the 15th International Congress of Phonetic Sciences (Barcelona). We express our thanks to the following: (a) Linda Hunsaker, for preparing the stimulus pictures; (b) Elaine Russo Hitchcock, for assistance in subject recruitment and recording; (c) several students from Long Island University, Brooklyn and New York University, for assistance in recording and data analysis: in alphabetical order, Giridhar Athmakuri, Ingrid Davidowitz, Linda Greenwald, Karen Keung, and Gabrielle Rothman; (d) Elliot Saltzman, for raising the issue of possible artifact in longer signals; and (e) Julia Irwin, Anders Löfqvist, Martha Tyrone, Brad Story, and 3 anonymous reviewers for helpful comments on earlier versions of this paper. This work was supported by NIH Grant No. DC-04473-03 to Haskins Laboratories and by MCT/CNPq (Brazil).

¹To our knowledge, the only other study that has examined variability using aerodynamic data is Zajac and Hackett (2002). Those authors assessed the timing of flow and pressure peaks in the /mp/ sequence of the word "hammer." Although aerodynamic signals were used, the aims and results of that study were more comparable to studies that have quantified temporal variability using acoustic signals than to the work presented here.

²Fewer than 15 productions were available in five cases (out of 90: 3 groups × 10 speakers × 3 consonants). All of these were in the 5 year olds, and three were from one child who produced 10–14 usable tokens each of /h s z/. Most of her /h/ exclusions reflected pausing; /s z/ exclusions were mostly due to voicing errors because she tended to perseverate when changing from one production to the next. Her data were not un-

usual in other respects; for example, her variability indices were not notably different from those of other 5 year olds.

³The number of splines was also selected by considering the expansion as a low-pass filtering process (Unser *et al.*, 1993a, 1993b). According to such theory, letting F be the sampling frequency of the tokens, n the number of samples in the tokens, N the number of spline functions and O their order, then the cutoff frequency f of the filtering process is $f = (F/2) * (N - O + 2) / n$. In the present case, $N = 42$, $O = 4$, and $F = 10$ kHz, and assuming $n = 5000$ points (approximate average number of samples of the tokens), we obtain $f = 40$ Hz. This frequency range is sufficient to yield a good representation of record shapes.

- Berthier, N. E. (1996). "Learning to reach: A mathematical model," *Dev. Psychol.* **32**, 811–823.
- Cheng, H. Y., Murdoch, B. E., Goozée, J. V., and Scott, D. (2007). "Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults," *J. Speech Lang. Hear. Res.* **50**, 375–392.
- Chermak, G. D., and Schneiderman, C. R. (1986). "Speech timing variability of children and adults," *J. Phonetics* **13**, 477–480.
- Deutsch, K. M., and Newell, K. M. (2001). "Age differences in noise and variability of isometric force production," *J. Exp. Child Psychol.* **80**, 392–408.
- Deutsch, K. M., and Newell, K. M. (2003). "Deterministic and stochastic processes in children's isometric force variability," *Dev. Psychobiol.* **43**, 335–345.
- Deutsch, K. M., and Newell, K. M. (2004). "Changes in the structure of children's isometric force variability with practice," *J. Exp. Child Psychol.* **88**, 319–333.
- Eguchi, S., and Hirsh, I. J. (1969). "Development of speech sounds in children," *Acta Oto-Laryngol.* **57**, 1–51.
- Ferguson, C. A., and Farwell, C. B. (1975). "Words and sounds in early language acquisition," *Language* **51**, 419–439.
- Goffman, L., Ertmer, D. J., and Erdle, C. (2002). "Changes in speech production in a child with a cochlear implant: Acoustic and kinematic evidence," *J. Speech Lang. Hear. Res.* **45**, 891–901.
- Green, J. R., Moore, C. A., and Reilly, K. J. (2002). "The sequential development of jaw and lip control for speech," *J. Speech Lang. Hear. Res.* **45**, 66–79.
- Hatze, H. (1986). "Motion variability: Its definition, quantification, and origin," *J. Motor Behav.* **18**, 5–16.
- Howe, M. S., and McGowan, R. S. (2005). "Aeroacoustics of [s]," *Proc. R. Soc. London, Ser. A* **461**, 1005–1028.
- Hughes, D., McGillivray, L., and Schmidek, M. (1997). *Guide to Narrative Language: Procedures for Assessment* (Thinking, Eau Claire, WI).
- Jensen, J. L., Thelen, E., Ulrich, B. D., Schneider, K., and Zernicke, R. F. (1995). "Adaptive dynamics of the leg movement patterns of human infants: III. Age-related differences in limb control," *J. Motor Behav.* **27**, 366–374.
- Johnson, K., Ladefoged, P., and Lindau, M. (1993). "Individual differences in vowel production," *J. Acoust. Soc. Am.* **94**, 701–714.
- Kent, R. D. (1976). "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," *J. Speech Hear. Res.* **19**, 421–447.
- Kent, R. D. (1992). "The biology of phonological development," in *Phonological Development: Models, Research, Implications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 65–90.
- Kent, R. D., and Forner, L. L. (1980). "Speech segment durations in sentence recitations by children and adults," *J. Phonetics* **8**, 157–168.
- Kewley-Port, D., and Preston, M. (1974). "Early apical stop production: A voice onset time analysis," *J. Phonetics* **2**, 195–210.
- Klatt, D. H., Stevens, K. N., and Mead, J. (1968). "Studies of articulatory activity and airflow during speech," in *Sound Production in Man*, edited by A. Bouhuys (New York Academy of Sciences, New York).
- Koenig, L. L. (2000). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," *J. Speech Lang. Hear. Res.* **43**, 1211–1228.
- Koenig, L. L., Mencl, W. E., and Lucero, J. C. (2005). "Multidimensional analyses of voicing offsets and onsets in female speakers," *J. Acoust. Soc. Am.* **118**, 2535–2550.
- Körding, K. P., and Wolpert, D. M. (2004). "Bayesian integration in sensorimotor learning," *Nature (London)* **427**, 244–247.
- Lecours, A. R. (1975). "Myelogenetic correlates of the development of speech and language," in *Foundations of Language Development: A Mul-*

- tidisciplinary Approach*, edited by E. H. Lenneberg and E. Lenneberg (Academic, New York), Vol. 1, pp. 121–135.
- Lee, S., Byrd, D., and Krivokapic, J. (2006). "Functional data analysis of prosodic effects on articulatory timing," *J. Acoust. Soc. Am.* **119**, 1661–1671.
- Lee, S., Potamianos, A., and Naryanan, S. (1999). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.* **105**, 1455–1468.
- Lisker, L., Abramson, A. S., Cooper, F. S., and Schvey, M. H. (1969). "Transillumination of the larynx in running speech," *J. Acoust. Soc. Am.* **45**, 1544–1546.
- Liu, Y.-T., Mayer-Kress, G., and Newell, K. M. (2006). "Qualitative and quantitative change in the dynamics of motor learning," *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 380–393.
- Lucero, J. C. (1999). "Computation of the harmonics-to-noise ratio of a voice signal using a functional data analysis algorithm," *J. Sound Vib.* **222**, 512–520.
- Lucero, J. C. (2005). "Comparison of measures of variability of speech movement trajectories using synthetic records," *J. Speech Lang. Hear. Res.* **48**, 336–344.
- Lucero, J. C., and Löfqvist, A. (2005). "Measures of articulatory variability in VCV sequences," *ARLO* **6**, 80–84.
- Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (1997). "On the registration of time and the patterning of speech movements," *J. Speech Lang. Hear. Res.* **40**, 1111–1117.
- Macken, M. A., and Barton, D. (1980). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," *J. Child Lang* **7**, 41–74.
- Miller, J. F., and Chapman, R. S. (1981). "Research note: The relation between age and mean length of utterance in morphemes," *J. Speech Hear. Res.* **24**, 154–161.
- Munson, B. (2004). "Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean," *J. Speech Lang. Hear. Res.* **47**, 58–69.
- Netsell, R. (1981). "The acquisition of speech motor control: A perspective with directions for research," in *Language Behavior in Infancy and Early Childhood*, edited by Rachel E. Stark (Elsevier, New York), pp. 127–156.
- Nittrouer, S. (1993). "The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study," *J. Speech Hear. Res.* **36**(5), 959–972.
- Nittrouer, S., and Studdert-Kennedy, M. (1987). "The role of coarticulatory effects in the perception of fricatives by children and adults," *J. Speech Hear. Res.* **30**, 319–329.
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," *J. Speech Hear. Res.* **32**, 120–132.
- Ohala, J. (1975). "The temporal regulation of speech," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 431–453.
- Ohde, R. N. (1985). "Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children," *J. Acoust. Soc. Am.* **78**, 1554–1561.
- Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (1996). "Functional data analyses of lip motion," *J. Acoust. Soc. Am.* **99**, 3707–3717.
- Ramsay, J. O., and Silverman, B. W. (1997). *Functional Data Analysis* (Springer-Verlag, New York).
- Ramsay, J. O., and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies* (Springer-Verlag, New York).
- Raphael, L. J., and Bell-Berti, F. (1975). "Tongue musculature and the feature of tension in English vowels," *Phonetica* **32**, 61–73.
- Sander, E. K. (1972). "When are speech sounds learned?," *J. Speech Hear. Disord.* **37**, 55–63.
- Sawashima, M. (1970). "Glottal adjustments for English obstruents," Haskins Laboratories Status Report on Speech Research **21/22**, 187–199.
- Schulz, G. M., Sulc, S., Leon, S., and Gilligan, G. (2000). "Speech motor learning in Parkinson Disease," *J. Med. Speech-Lg. Path.* **8**, 243–247.
- Scully, C., Castelli, E., Brearley, E., and Shirt, M. (1992). "Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives," *J. Phonetics* **20**, 39–51.
- Shadle, C. (1990). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 187–209.
- Sharkey, S. G., and Folkins, J. W. (1985). "Variability in lip and jaw movements in children and adults: Implications for the development for speech motor control," *J. Speech Hear. Res.* **28**, 8–15.
- Smith, A., and Goffman, L. (1998). "Stability and patterning of speech movement sequence in children and adults," *J. Speech Lang. Hear. Res.* **41**, 18–30.
- Smith, B. L. (1978). "Temporal aspects of English speech production: A developmental perspective," *J. Phonetics* **6**, 37–67.
- Smith, B. L. (1992). "Relationships between duration and temporal variability in children's speech," *J. Acoust. Soc. Am.* **91**, 2165–2174.
- Smith, B. L. (1994). "Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech," *J. Phonetics* **22**, 155–175.
- Smith, B. L. (1995). "Variability of lip and jaw movements in the speech of children and adults," *Phonetica* **52**, 307–316.
- Smith, B. L., Kenney, M. K., and Hussain, S. (1996). "A longitudinal investigation of duration and temporal variability in children's speech production," *J. Acoust. Soc. Am.* **99**, 2344–2349.
- Smith, B. L., and McLean-Muse, A. (1986). "Articulatory movement characteristics of labial consonant productions by children and adults," *J. Acoust. Soc. Am.* **80**(5), 1321–1328.
- Smith, B. L., Sugarman, M. D., and Long, S. H. (1983). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," *J. Acoust. Soc. Am.* **74**, 744–749.
- Sosnik, R., Hauptmann, B., Karni, A., and Flash, T. (2004). "When practice leads to coarticulation: The evolution of geometrically defined movement primitives," *Exp. Brain Res.* **156**, 422–438.
- Stathopoulos, E. T. (1995). "Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech," *J. Phonetics* **23**, 67–80.
- Thelen, E., Corbetta, D., and Spencer, J. P. (1996). "Development of reaching during the first year: Role of movement speed," *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1059–1076.
- Tingley, B. M., and Allen, G. D. (1975). "Development of speech timing control in children," *Child Dev.* **46**, 186–194.
- Todorov, E., and Jordan, M. I. (2002). "Optimal feedback control as a theory of motor coordination," *Nat. Neurosci.* **5**, 1226–1235.
- Unser, M., Aldroubi, A., and Eden, M. (1993a). "B-spline signal processing: Part I—Theory," *IEEE Trans. Signal Process.* **41**, 821–833.
- Unser, M., Aldroubi, A., and Eden, M. (1993b). "B-spline signal processing: Part II—Efficient design and applications," *IEEE Trans. Signal Process.* **41**, 834–848.
- Vetter, P., and Wolpert, D. M. (2000). "The CNS updates its context estimate in the absence of feedback," *NeuroReport* **11**, 3783–3786.
- Vihman, M. M. (1993). "Vocal motor schemes, variation and the perception-production link," *J. Phonetics* **21**, 163–169.
- von Hofsten, C. (1989). "Motor development as the development of systems: Comments on the Special Section," *Dev. Psychol.* **25**, 950–953.
- Walsh, B., and Smith, A. (2002). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," *J. Speech Lang. Hear. Res.* **45**, 1119–1133.
- Waterson, N. (1971). "Child phonology: A prosodic view," *J. Linguistics* **7**, 179–211.
- Yan, J. H., Stelmach, G. E., Thomas, J. R., and Thomas, K. T. (2000). "Developmental features of rapid aiming arm movements across the lifespan," *J. Motor Behav.* **32**, 121–140.
- Zajac, D. J., and Hackett, A. (2002). "Temporal characteristics of aerodynamic segments in the speech of children and adults," *Cleft Palate Craniofac J.* **39**, 432–438.