

Studying articulatory variability using Functional Data Analysis

Laura L. Koenig^{*}, Jorge C. Lucero,[†] and Anders Löfqvist[‡]

^{*} Haskins Laboratories & Long Island University, Brooklyn Campus

[†] University of Brasilia

[‡] Haskins Laboratories & Lund University

E-mail: koenig@haskins.yale.edu, lucero@mat.unb.br, lofqvist@haskins.yale.edu

ABSTRACT

Production variability is frequently used as an index of the stability and maturity of the speech motor system. Here, we assess speech production variability using Functional Data Analysis (FDA). In FDA, a smooth warping function of time is used to align the data, and separate indices of temporal and amplitude variability are obtained. Results on adult kinematic data show that amplitude variability across articulators varies with place of constriction, suggesting task-specific production constraints. Amplitude and phasing indices for children and adults indicate both group differences (higher average indices in children relative to adults) and individual patterns of variability. Pilot data from a child treated for apraxia of speech show variability indices within normal limits. Overall, our work suggests that FDA may be useful for quantifying variability across populations and speaking tasks. This may have clinical utility in differentiating atypical speech production behavior, and/or in charting the success of treatment.

1. INTRODUCTION

Many researchers have used measures of token-to-token variability to compare child and adult speech [e.g., 1–4]), and normal and clinical populations [e.g., 5–6]. These studies generally indicate that elevated variability characterizes both immature (or developing) speakers and those with breakdown of the speech production system. At the same time, the nature and extent of variability in normal adult speakers is not fully known. Current theories of motor control [7] suggest that knowing the regions of variability and relative stability in repeated motor acts may provide insight into speakers' control systems. The goals of our work here are (a) to determine how variability in adult speakers differs across articulators during consonant production; (b) to investigate whether measures of temporal and amplitude variability differentiate children and adults; and (c) to see how normal and clinical populations differ in their variability profiles. To address these questions, we use the techniques of Functional Data Analysis (FDA), a set of procedures designed for data that may be considered functions of time [8]. We extend the application of FDA to data types and populations not

considered in past work [9–11]. Although the specifics of the FDA procedures differ somewhat for the two data sets presented below, the analyses share the general feature of time-warping the signals so as to bring their events into closer temporal alignment.

2. METHODS

Adult kinematic data: Four normal adults produced 10 repetitions of VCVs in the phrase "Say ___ again." Analysis here focuses on the sequences /api, ata, aka/. Movement signals were digitized at 625Hz using a magnetometer, with receivers on the jaw, upper and lower lips, and at four locations on the tongue: One on the tip, one as far back as possible, and the other two spaced evenly between the front and back extremes. These four locations, from anterior to posterior, will be referred to as tongue tip (TT), tongue blade (TBL), tongue body (TB), and tongue root (TR). A receiver on the nose bridge was used to correct for head movement, and the data were rotated to align the x-axis with the occlusal plane. For the jaw and lip receivers, only displacements in y were considered as movement in the x-axis is typically minimal [12]. For the tongue signals, the movement traces were decomposed into x and y components.

Child aerodynamic data: Normal 4–5 and 9–10 year old English-speaking children were recorded producing 25–30 repetitions of intervocalic /h, s, z/ in the utterances "Papa Hopper/Sapper/Zapper." A 10-year-old child formerly treated for developmental apraxia of speech (DAS) was also recorded. Mothers of the children were recorded to provide adult data for comparison. Oral airflow signals were collected using a Rothenberg mask. Signals were digitized at 10kHz on a PowerLab recording system. The data were subsequently smoothed to remove glottal pulsing, and the DC flow trajectories were used to assess laryngeal abduction for /h/, tongue-tip constriction for /z/, and combined laryngeal-oral gestures for /s/.

FDA: Kinematic signals The kinematic signals for each consonant were read into Matlab, vertically aligned by subtracting out the mean of the set, resampled to a common length of 201 points, and put into vector form:

$$x_i(t) = (TTY_i(t), TTX_i(t), TBLY_i(t), \dots, JAWY_i(t)) \quad (1)$$

where $i=1, \dots, N$, with N =the number of repetitions. Time-warping functions were generated to minimize the following measure of proximity between the original

records and their average [8]:

$$C = \sum_{i=1}^N \sum_{j=1}^M \left\{ \alpha_j \int_0^1 [x_{ij}^*(t) - \bar{x}_j^*(t)]^2 dt + \beta_j \int_0^1 [Dx_{ij}^*(t) - D\bar{x}_j^*(t)]^2 dt + \lambda \int_0^1 w_j^2 dt \right\}$$

where

$$\alpha_j = \frac{1}{\int_0^1 \bar{x}_j^2(t) dt} \quad \beta_j = \frac{1}{\int_0^1 D\bar{x}_j^2(t) dt} \quad (2)$$

M is the number of recorded signals, and λ is a roughness penalty coefficient imposed on the curvature w of the warping function. Applying the warping functions to the data has the effect of bringing regions of displacement in all signals into closer alignment. The effect for a single articulator is shown in Figure 1. Because the derivatives have more events (or variability) than the original signals, their inclusion in this analysis provides better alignment than if the kinematic signals alone were used in the cost function. Finally, an index of amplitude variability was calculated over a region defined by peaks in the first derivative corresponding to the movements into and out of the consonantal closure by the active articulator.

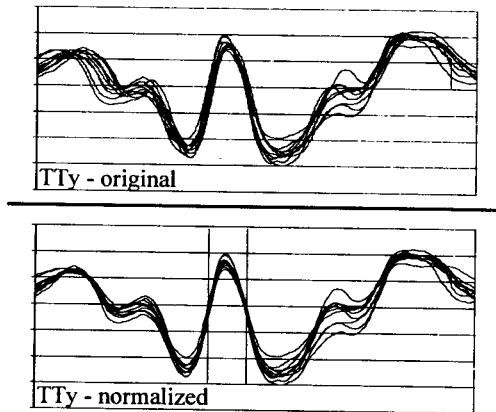


Figure 1: Kinematic records for one speaker producing /ata/, before and after normalization. The vertical lines in the bottom plot indicate the region over which the amplitude index was calculated.

FDA: Aerodynamic signals: The aerodynamic data were down-sampled by a factor of 10 to facilitate Matlab processing. The records were trimmed to begin and end with peaks in the second derivative corresponding to the flanking /p/ releases and closures. Each record was amplitude-normalized by subtracting the mean and dividing by the standard deviation. Time-warping functions were generated to minimize a cost function calculated on the DC flow signals:

$$C = \sum_{i=1}^N \int_0^1 [x_i^*(t) - \bar{x}^*(t)]^2 dt + \lambda \int_0^1 w_i^2(t) dt \quad (3)$$

Because the aerodynamic signals showed considerable

variability (especially for the children), the first derivatives of the signals were not included here as they were for the adult analysis. The standard deviation of the warping functions for a given subject and consonant was used to measure temporal (or phasing) variability. Amplitude variability indices were computed as the standard deviation over regions defined by minima flanking (a) the flow peak for /h/ and (b) the two flow peaks bordering the tongue-tip constriction for /s/ (see Figure 2).

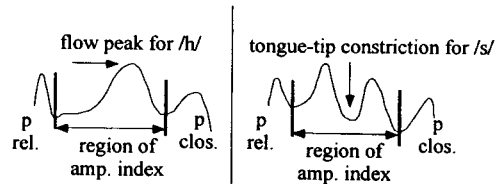


Figure 2: Sample aerodynamic signals of /h/ (left) and /s/ (right), showing region over which amplitude index was calculated.

3. RESULTS

Adults: Qualitative inspection of the aligned data for the lingual consonants /t, k/ suggested that, compared to the rest of the utterance, variability was low during the consonantal opening and closing movements in the portion of the tongue involved in making the closure (tongue tip and tongue root, respectively). Vertical variation was also more limited than horizontal.

The amplitude indices for all articulators are shown in Figure 3 for the utterances /aka/ (left), /ata/ (middle) and /api/ (right). The data points representing the active articulators are indicated with filled arrows: For /aka/, TRy; for /ata/, TTy; and for /api/, LLy. The receiver on the jaw is also shown (unfilled arrows) for /ata/ and /api/. (Jaw contributions are limited in /k/ closures compared to more anterior constrictions [e.g., 13]).

Figure 3a shows that, during the /k/ closure, vertical variability in the tongue root (TRy) is relatively low compared to other articulators. This suggests that movements of the posterior tongue are tightly constrained during production of velar closures. In 3b, variability in the jaw is among the lowest of all articulators during the alveolar closure for /t/. The indices for the tongue tip are higher than those for the jaw for all speakers. For 3 of the 4 speakers, tongue tip variability is relatively low, but this is not true for S1. The low variability indices for the jaw are consistent with past reports indicating that the jaw is a major contributor to raising and anchoring the tongue for alveolars [e.g., 14].

The results for the bilabial stop /p/ show a rather different pattern. Figure 3c indicates that variability for the active articulator (LLy) here is among the highest of all articulators. Variability in the jaw is somewhat

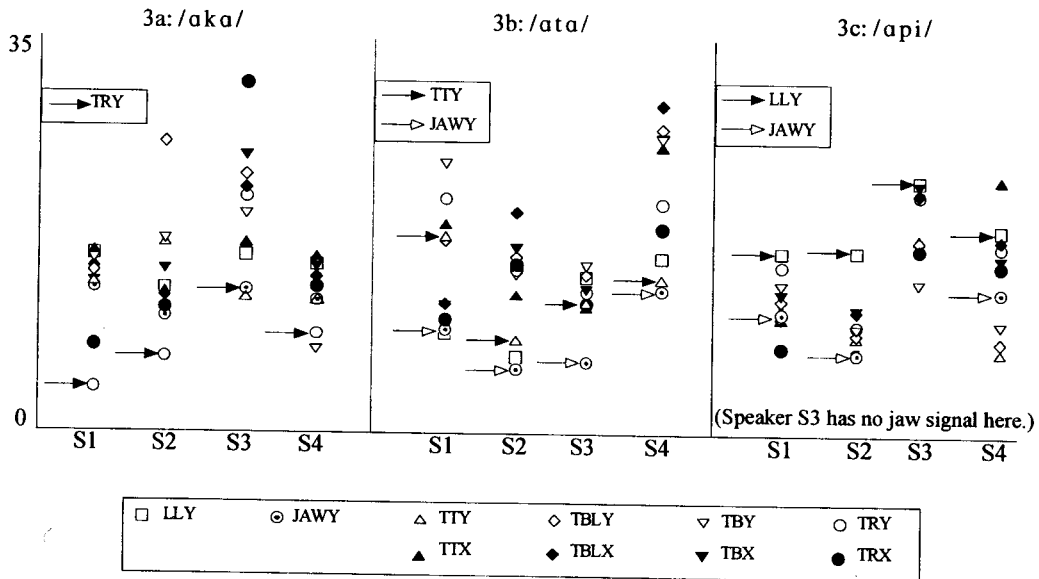
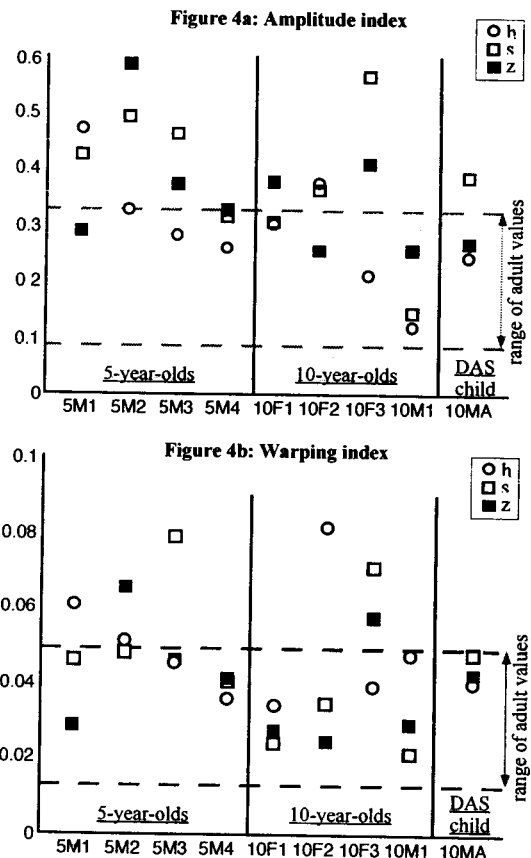


Figure 3: Amplitude variability indices for adult kinematic data. The filled arrows indicate the active articulator for forming the medial consonant constriction. Unfilled arrows indicate the jaw.

lower than that for the lower lip, but is not necessarily low relative to other articulators. We hypothesize that, in bilabial closures, tissue compression upon lip contact may yield higher variability than in cases of an articulator meeting a rigid surface such as the hard palate.

In sum, we conclude that articulatory variability in normal adults varies as a function of both the phonetic requirements of the consonant and the biomechanical characteristics of the articulatory structures involved.

Children: Amplitude and warping indices for the child speakers are shown in Figure 4. For reference, the ranges of values produced by 4 adult female speakers are indicated by dotted lines. The child treated for DAS is at the right of each figure. For both the amplitude and the warping indices, the children's values are, on average, higher than the adults', but some children fall into the adult range on both measures. Overlap with the adult range is more common among the 10-year-olds than the 5-year-olds. These results support past work showing higher variability among children than adults, with a developmental trend towards more adult-like values into the pubertal years, but extensive cross-subject variability [1-4]. For a given child, consonantal patterns of amplitude and warping variability tend to be similar, but there are some exceptions (e.g., 10M1). For the data shown here, the correlation between amplitude and warping indices is significant at $p=.0002$, but the R^2 value is only .41. Thus, amplitude and temporal variability may show somewhat different profiles in individual speakers. This observation also holds for the adult speakers we have analyzed.



Figs. 4. Amplitude (top) and warping (bottom) indices for 5- and 10-year-old children and for a 10-year-old treated for developmental apraxia of speech. The range of adult values in each figure is based on results from 4 adult females.

One characteristic typically ascribed to DAS is variability of speech output, thought to reflect a deficit in spatiotemporal control [cf. 15]. Based on this, we had anticipated that the 10-year-old treated for DAS might show elevated production variability relative to his normally-developing peers, even after completing a 4-year treatment program. Figure 4 shows that this is not the case: subject 10MA's indices are well within the range for his age. These results are consistent with his current clinical profile of normal articulation despite persistent subtle language difficulties (syntax, semantics, pragmatics). In future work, we plan to explore how FDA indices vary longitudinally in children being treated for DAS. Measures of variability such as those provided by FDA may provide a method for assessing speech production characteristics of persons with motor speech disorders, and for charting treatment progress.

4. CONCLUSIONS

We believe that use of FDA to quantify speech production variability has wide applicability in studying questions of speech motor control. Our data on normal adult speakers indicate differences in variability measures across articulators which can be interpreted both in terms of the linguistic constraints on speech movements and the biomechanical characteristics of the articulators in question. Comparisons between child and adult speakers show similar results as past work, but further indicate that temporal and amplitude variability are somewhat independent within speakers. Obtaining these separate measures may provide greater insight into the nature of variability in speech development. Finally, these methods may have clinical utility by providing detailed profiles of an individual speaker's production variability, which could be used to assess the efficacy of treatment.

5. ACKNOWLEDGMENTS

This work was supported by NIH grants to Haskins Laboratories (DC 00865 and DC 04473-01) and by CNPq, Brazil. We thank our subjects and the parents of the children for their participation. We are grateful to Elaine Hitchcock for providing the clinical history of the child with DAS, and for assistance in child subject recruitment.

REFERENCES

- [1] S. Eguchi and I. J. Hirsh, "Development of speech sounds in children," *Acta Otolaryngologica*, vol. 57, 1969.
- [2] S. G. Sharkey, and J. W. Folkins, "Variability in lip and jaw movements in children and adults: Implications for the development for speech motor control," *Journal of Speech and Hearing Research*, vol. 28, pp. 8–15, 1985.
- [3] A. Smith and L. Goffman, "Stability and patterning of speech movement sequences in children and adults," *Journal of Speech, Language, and Hearing Research*, vol. 41, pp. 18–30, 1998.
- [4] B. L. Smith, "Variability of lip and jaw movements in the speech of children and adults," *Phonetica*, vol. 52, pp. 307–16, 1995.
- [5] K. S. Harris, J. Rubin-Spitz, and N. S. McGarr, "The role of production variability in normal and deviant developing speech," *ASHA Reports 15*, pp. 50–57. Rockville, Maryland, 1985.
- [6] G. Weismer, R. Martin, R. D. Kent, and J. F. Kent, "Formant trajectory characteristics of males with amyotrophic lateral sclerosis," *Journal of the Acoustical Society of America*, vol. 91, pp. 1085–98, 1992.
- [7] E. Thelen and L. B. Smith, *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press/Bradford, 1994.
- [8] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. New York: Springer-Verlag, 1997.
- [9] J. O. Ramsay, K. G. Munhall, V. L. Gracco, and D. J. Ostry, "Functional data analyses of lip motion," *Journal of the Acoustical Society of America*, vol. 99, pp. 3707–17, 1996.
- [10] J. C. Lucero, K. G. Munhall, V. L. Gracco, and J. O. Ramsay, "On the registration of time and the patterning of speech movements," *Journal of Speech, Language, and Hearing Research*, vol. 40, pp. 1111–17, 1997.
- [11] J. Lucero, and L. L. Koenig, "Time normalization of voice signals using functional data analysis," *Journal of the Acoustical Society of America*, vol. 108, pp. 1408–20, 2000.
- [12] A. Löfqvist, and V. L. Gracco, "Lip and jaw kinematics in bilabial stop consonant production," *Journal of Speech, Language, and Hearing Research*, vol. 40, pp. 877–93, 1997.
- [13] R. D. Kent, and K. L. Moll, "Cinefluorographic analyses of selected lingual consonants," *Journal of Speech and Hearing Research*, vol. 15, pp. 453–73, 1972.
- [14] M. Stone, and E. Vatikiotis-Bateson, "Trade-offs in tongue, jaw, and palate contributions to speech production," *Journal of Phonetics*, vol. 23, pp. 81–100, 1995.
- [15] L. Shriberg, D. M. Aram, and K. Joan, "Developmental apraxia of speech: I. Descriptive and theoretical perspectives," *Journal of Speech, Language, and Hearing Research*, vol. 40, pp. 273–85, 1997.