

## CHARACTERISTICS OF SPEECH MOVEMENTS

Vincent L. Gracco<sup>1</sup>, Anders Löfqvist<sup>1</sup>, James O. Ramsay<sup>2</sup>, Kevin G. Munhall<sup>3</sup>, & David J. Ostry<sup>2</sup>

<sup>1</sup>Haskins Laboratories, New Haven, CT; USA<sup>2</sup> Department of Psychology, McGill University, Montreal, Canada; <sup>3</sup> Department of Psychology, Queen's University, Kingston, Ontario, Canada

### RESUME

Pour approfondir notre connaissance de la production de la parole et ses mécanismes sensori-moteurs, il faut utiliser des approches multiples et diversifiées. Cet article s'occupe des méthodes pour étudier les facteurs spatiaux et temporels de la production de la parole avec le but de mieux comprendre ses principes fondamentaux. Nos données mettent au point la précision nécessaire pour des mouvements articulatoires différents, ainsi que comment le degré de cette précision est gouverné par le contexte sur le plan segmental. Nous proposons aussi une méthode pour étudier les variations cinématiques temporelles des mouvements articulatoires. Nos résultats mettent en relief la nature complexe et dynamique de la production de la parole.

### ABSTRACT

In order to more completely understand speech production and the underlying sensorimotor control, a number of complementary approaches are needed. The following focuses on methods for examining the spatiotemporal dynamics of speech production to infer characteristics of the underlying neural control processes. Data are presented on the precision with which different speech articulators are controlled, the variation in control precision as a function of phonetic context, and a method for examining kinematic variation as a function of time. Results highlight the dynamic multivariate nature of the speech production process.

### INTRODUCTION

Speech production is a complex process involving time-varying neural commands and a constantly changing physiological environment. For speech, the overarching goal is to communicate, and accurate information transfer is based on a balance between the perceptual and production systems. One method that can be used to understand the balance between these two

processes generally, and the speech production process specifically, is to examine the output variability under a variety of experimental conditions. One goal of speech research is to identify the sources of variation in an attempt to understand the link between these two important processes.

Determining the extent and degree of articulatory control and the associated variability, however, is not a simple task. Articulatory variations occur in the spatial positioning of various vocal tract structures during consonant and vowel production, in the paths that different articulators take to and from constriction locations, and in the so-called motor equivalence trade-offs among goal-related articulators. In addition, the duration of linguistically significant events that unfold in the vocal tract range from less than 100 milliseconds to many seconds. As such, the character of the system processes operating to produce linguistically meaningful information may vary substantially as a function of time.

Moreover, quantitative evaluation of speech movement variability is often undertaken from a reduced description of the relevant data. For example, speech movement analyses are often restricted to simple summaries of the observable motion in a single spatial dimension at a single point in time. In order to completely understand the speech production process an approach is needed that allows for examination of point estimates along with richer descriptions of articulator configurations and the associated time histories of articulator motion. With regard to the latter issue, one such approach is that offered by Ramsay and colleagues (Ramsay, 1982; Besse & Ramsay, 1986) in which traditional multivariate statistical analyses are applied to continuous motion and expressed in functional analytic terms. Functional data analysis (FDA) has been shown to deal quantitatively with multidimensional time series data during

simple tongue (Ramsay, 1982; Besse & Ramsay, 1986) and lip motion (Ramsay, Munhall, Gracco, & Ostry, 1996).

In the following, both two-dimensional spatial point measurements, tongue surface estimations, and functional analytic techniques are applied to a corpus of data to obtain a richer understanding of tongue kinematics during a variety of speech tasks.

## METHODS

The data were obtained from two different experiments, the first of which employed four subjects (two males, two females). For this experiment, subjects produced a number of words in the carrier phrase "Say \_\_\_\_\_ again." and repeated each ten times. This corpus consisted on mono- and bi-syllabic words with the consonant /r/, /s/, /l/, or /n/ in the initial position and /k/, /g/, /t/, and /d/ syllable-finally. The second experiment examined continuous tongue motion variation over multiple repetitions of a single multisyllabic word. For this experiment, three subjects (one male, two females) repeated the phrase "The oleander is a flower" 35 times with the word "oleander" as the target for analysis.

Movements of the mandible and four points on the tongue surface were obtained using an electromagnetic transduction device (Perkell et al., 1992). Receivers were placed in the mid-sagittal plane of the device at these anatomical locations: the bridge of the nose, the maxillary gum line, the mandibular gum line, and four locations on the tongue (from front to back [TT], [TBL], [TB], [TR]). The maxillary and nose receivers were used to correct for any head motion. The tongue receivers were placed approximately 1 cm behind the tongue tip and spaced on average 1 cm apart. Estimates of the tongue surface were made from cubic spline interpolations of the average receiver locations obtained at selected acoustic time points.

All movement signals were low pass filtered (200 Hz) in hardware and digitally sampled at 625 Hz (12 bit resolution). Following the digitization, the voltages were smoothed (36.8 ms triangular window) and converted to positions in the midsagittal plane of the device. All data were corrected for head motion and rotated to the subjects' occlusal plane.

Two dimensional spatial variation associated with the initial phonetic segments /r/, /s/, /l/, and /n/ was examined

for the jaw receiver and four tongue receivers at the acoustically-defined onsets. For the voiceless sound /s/, the onset of the friction noise was identified. The sagittal plane coordinates of each receiver were obtained at this time point. For the voiced sounds /r/, /l/, and /n/, the same procedure was used except that acoustic onset was used rather than the onset of friction.

## RESULTS

### Articulator Variation

For both data sets, the degree of two-dimensional positional variability was found to be articulator-dependent. For all subjects, the mandible displayed less variation than any of the four tongue receivers. Figure 1 shows data from the first experiment consonants contrasting the tongue and jaw receiver variations.

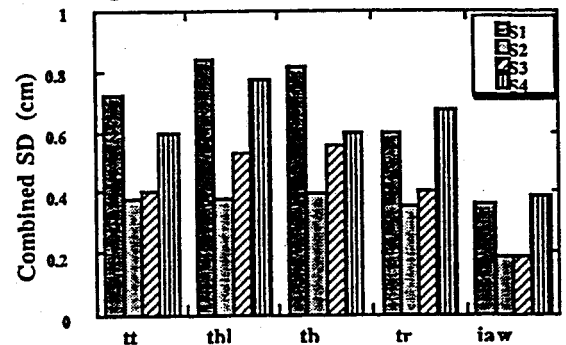


Figure 1. One standard deviation collapsed for the two spatial dimensions (X plus Y) for each receiver location expressed in cm. for the four subjects.

The combined measure of variation reflects the sum of the standard deviation of the mean X and Y positions obtained at each acoustically-defined measurement point. As can be seen, the variation is always less for the jaw receiver compared to the tongue receivers.

Comparison of the average jaw and tongue front and rear receiver positions also demonstrates more contextual spatial variation for the tongue than the jaw (Figure 2).

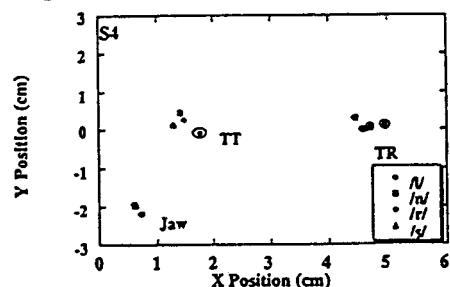


Figure 2. Average JAW and tongue tip (TT) and tongue rear (TR) receiver positions (and one

standard deviation ellipses) for the four different initial consonants for a single subject.

**Effects of Phonetic Identity**

The degree of variation in the spatial positioning within and across consonants and vowels can be used to infer the degree of control precision or sensitivity of the control process to the underlying speech movement goals. Although the corpus examined in the first experiment was limited in scope, there was a tendency for phonetic identity to be a factor in the degree of tongue position variation. As illustrated in Figure 3, receiver positions for /r/ were more variable than for any of the other phones examined while the receiver positions for /n/ were the least variable.

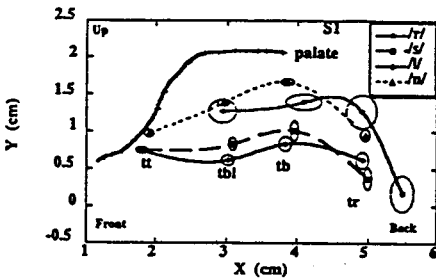


Figure 3. Tongue surface estimations for the four different initial consonants. from a single subject (N=10) Shown are the average receiver locations and one standard deviation ellipses.

It should be noted, however, that while the same pattern of variability was consistent across subjects for word-initial consonants, different results were observed when the context was changed.

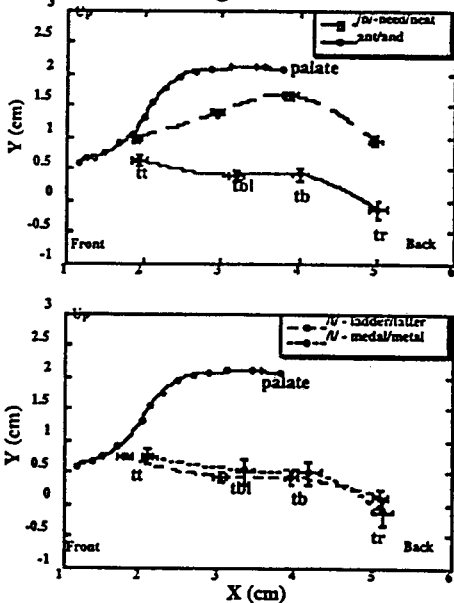


Figure 4. Average tongue surface estimates for a single subject (N=10) comparing /n/ syllable initial

versus syllable medial (top) and /l/ syllable initial versus syllable final (bottom).

The tongue surface shape associated with the word initial consonant /n/ was found to be stable, as evidenced by small positional variance across the ten repetitions. However, when /n/ was produced syllable medially, the tongue surface was found to be quite different compared to the syllable initial tongue surface (top panel, Figure 4). This is not surprising since perceptual saliency is more heavily dependent on the presence of nasality than tongue shape. In contrast, the tongue surface for /l/ was found to be less variable across contexts suggesting that tongue shape is important for correct production/perception of /l/ (bottom panel, Figure 4).

A related observation was made for the vowel /ae/. The top portion of Figure 5

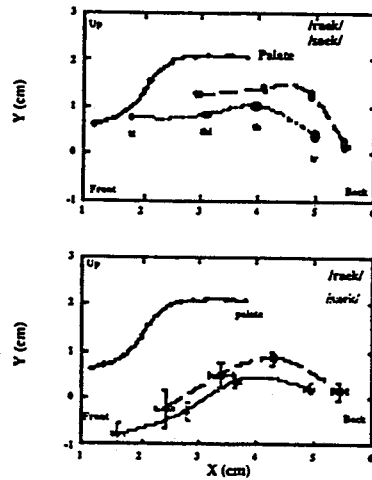


Figure 5. Average tongue surface estimates for /r/ and /s/ in the words "rack" and "sack" (top) and the tongue surface estimates for the vowel /ae/ in the respective contexts (bottom).

presents the tongue surface for /r/ and /s/ in the words "rack" and "sack". Shown in the bottom portion of the figure is the tongue surface for the vowel /ae/ in which the effects of the preceding consonant can be clearly seen. While the shape of the tongue surface for the vowel is similar across the two contexts, the actual positions of the receivers differ, apparently due to coarticulatory influences of the preceding consonant. It can also be seen that the amount of variation tolerated for /ae/ is quite large.

**Functional Analysis of Speech Motion**

A final analysis focused on examination of the variability of speech motion across

an interval spanning approximately 600 ms. For this analysis, multiple repetitions ( $n=35$ ) were obtained and temporally aligned based on tongue velocity associated with /o/ and /r/ in "oleander". The observations consisted of 10 movement streams (5 receivers, 2 spatial dimensions) for each of the 35 trials. The movement streams varied in duration from record to record with the number of sampled points ranging from 315 to 393 (approximately 500 to 629 ms). To simplify the analysis, the data were interpolated for each subject.

Shown in Figure 6 are the standard deviations of the jaw and tongue receiver positions for each spatial dimension as a function of utterance duration.

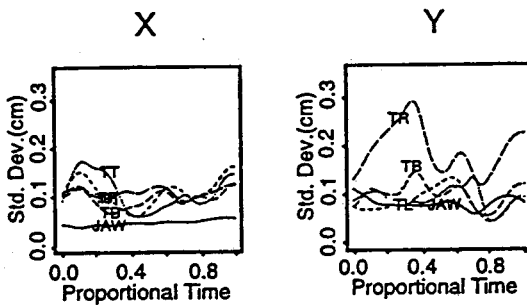


Figure 6. Average standard deviation (cm) as a function of time in the X and Y dimensions for the jaw receiver and the four tongue receivers

The jaw variation is relatively small and consistent across the utterance. The pattern of variation in the X dimension (front-back with respect to the occlusal plane) is relatively similar across the tongue receivers, while the pattern of variation in the Y dimension (up-down with respect to the occlusal plane) differs according to the location of the receiver. The rear tongue receivers (TR & TB) display a pattern of variation that is different than that found for the front of the tongue (TT & TBL). The rear receiver displays the greatest amount of variation with a peak in the function at approximately one-third of the way through the utterance corresponding roughly to the production of the /i/ in "oleander" and a smaller peak associated with the /r/.

FDA can also be used to partition the data into functional components of variation applying, for example, principal components analysis to movement functions. It is also possible to examine the derivatives of functions as well as the functions themselves. Figure 7 presents the eigenstructure of acceleration for each tongue receiver as a function of time. From

this display the different time course of the acceleration variance for each tongue receiver can be visualized.

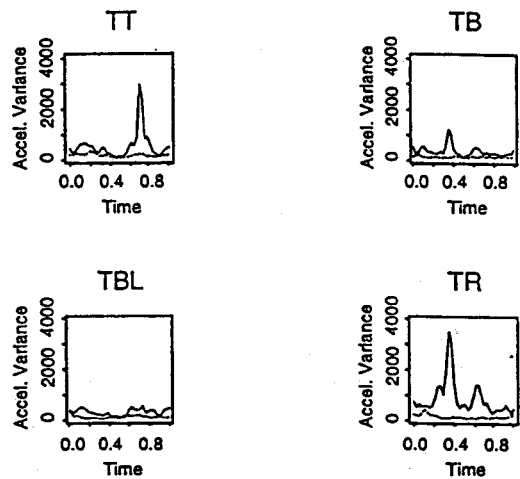


Figure 7. The eigenstructure of acceleration across time for the four tongue receivers.

## DISCUSSION

The present investigation was undertaken to examine speech movement variability using a number of complementary approaches. It was found that the degree of movement variation was related to articulator identity, phonetic identity, and phonetic context. In addition, a limited analysis of the dynamic nature of articulator motion found that variability across multiple tongue regions varies as a function of time. It is suggested that the multivariate and nonlinear nature of the speech production process requires the application of multiple kinds of analyses.

## ACKNOWLEDGMENT

This work was supported in part by research grant number DC-00865 and DC-00594 from the NIDCD, NIH.

## REFERENCES

- [1] Besse, P. & Ramsay, J.O. (1986). Principal components analysis of sampled functions. *Psychometrika*, 51, 285-311.
- [2] Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *JASA*, 92, 3078-3096.
- [3] Ramsay, J.O. (1982). When the data are functions. *Psychometrika*, 47, 379-396.
- [4] Ramsay, J.O., Munhall, K. G., Gracco, V.L., & Ostry, D.J. (in press, 1996). Functional data analysis of lip motion. *JASA*.