TR－H－113

Functional Data Analyses of Lip Motion

J.O. Ramsay (McGill University)
K.G. Munhall
V.L. Gracco (Haskins Laboratories)
D.J. Ostry (McGill University)

1994. 12. 12

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

Functional Data Analyses of Lip Motion

J.O. Ramsay[1]

K.G. Munhall[2]

V.L. Gracco[3]

D.J. Ostry[1]

Submitted to the Journal of the Acoustical Society of America

[1]McGill University, Montreal, Canada; [2]Queen's Unversity, Kingston, Canada and ATR Human Information Processing Research Laboratories, Kyoto, Japan; [3]Haskins Laboratories, New Haven , U.S.A.

Request for preprints should be directed to the first author at:

Department of Psychology
1205 Dr. Penfield Ave.
Montréal, Québec,
Canada H3A 1B1

electronic mail address: ramsay@psych.mcgill.ca

Abstract

The vocal tract's motion during speech is a complex patterning of the movement of many different articulators according to many different time functions. Understanding this myriad of gestures is important to a number of different disciplines including automatic speech recognition, speech and language pathologies, speech motor control and experimental phonetics. Central issues are the accurate description of the shape of the vocal tract and determining how each articulator contributes to this shape. A problem facing all of these research areas is how to cope with the multivariate data from speech production experiments.

In this paper we describe techniques that provide useful tools for describing multivariate functional data such as the measurement of speech movements. Our choice of data analysis procedures has been motivated by the need to partition the articulator movement in various ways: end-effects separated from shape effects, partitioning of syllable effects, and the splitting of within-ired variation from between-ired variation. The techniques of functional data analysis seem admirably suited to the analyses of phenomena such as these. Familiar multivariate procedures such as analysis of variance and principal components

analysis have their functional counterparts, and these reveal in a way more suited to the data the important sources of variation in lip motion. Finally, we found that the analyses of acceleration were especially revealing in considering the character of possible control mechanisms. Our focus is on using these speech production data to understand the basic principles of coordination. However, we believe the tools will have a more general use.

1 Introduction

Three major measurement problems have plagued speech scientists since the first recordings of articulation.The first is that the vocal tract is inaccessible to many simple measurement techniques. While many new recording technologies are available (e.g., EMMA, OPTOTRAK, X-ray Microbeam), this remains a serious problem. Secondly, the movements of the vocal tract are spatially complex and there is significant motion in 3-D. This leaves the researcher with data that have many dimensions and therefore many degrees of freedom. For soft-tissue articulators such as the lips and tongue, it is not clear how many dimensions are needed to adequately describe them. Some attempts have been made to assess the dimensionality of static lip (Linker, (1982) and tongue (Harshman, Ladefoged, & Goldstein, 1977) shapes but there has been no work that addresses the dimensionality of the motions of these articulators. The final problem is that the movement data from speech experiments arise from an underlying continuous process, but no account of this is taken in the analyses. The standard approach in speech production research is to digitally sample the movement trajectory and measure this sampled waveform at a small number of measurement points (peak velocity or maximum and minimum displacement). These small number of points greatly under-determine the variance in the trajectories and this approach takes no account of the functional variance in the data. In one sense, the sampled waveform is treated as if it was a set of independent measures rather than a sampled function.The present paper presents multivariate statistical techniques that address the latter two problems. We aim to demonstrate the utility of this approach by analyzing a set of lip movements, measured by positioning eight detectors around the upper and lower lips, and registering the position of each in three spatial coordinates over the time taken to utter a syllable. Thus this study is a prototype for experiments in which multi-dimensional movements of multiple articulator positions are recorded under various experimental conditions.

The statistical analysis of the data will use one of a family of techniques called functional data analysis (FDA). FDA involves the use of tools familiar in multivariate data analysis, such as principal components analysis, linear modeling, and canonical correlation analysis, to analyze data in which the basic observation is a function. Thus, instead of the multivariate analysis of a data matrix with values x_{ij} where $i = 1....N$ is the index of replications and $j = 1....p$ is the index of variables, FDA involves the analysis of replicated functions $x_i(t)$, where x_i varies continuously with values of its argument t. In the current study multiple functions x, corresponding to three spatial coordinates of eight separate articular positions observed under four experimental conditions, will be examined. The FDA techniques used in this paper all deal with a partition or decomposition of data into fundamental components of variation. Principal components analysis adapted to functional data serves the central objective of the paper, which is to assess the complexity and dimensionality of across-replication variation in lip movement, taken both within and across recording positions. FDA also permits the statistical analysis of derivatives of functions as well as the observed functions themselves. In this paper emphasis will also be placed on study of the second derivative of motion, since from physical principles one expects that the influence of forces (internal and external) have their most direct impact on acceleration and provide insight into the motor control process.

2 Method

2.1 Experimental Procedures

The subject was a male native speaker of Canadian English with no reported speech or language disorders. During the experiment the subject spoke CVC nonsense syllables in the carrier phrase "Say CVC again". The C in the utterance was /b/ and the vowels were / æ /, /i/, /u/, and /a/, or "bab", "beeb", "boob" and "bob". The subject spoke 20 repetitions of each syllable type, randomized across the 4 vowels.

The motion of the face was monitored using OPTOTRAK, an optoelectronic tracking system that can transduce the 3D position of markers. Eight infrared emitting diodes (ireds) were attached to the vermilion border of the lips using double-sided tape. Additional ireds were used to track the head during the experiment. This enabled us to correct for head movement and to transform lip motions to a coordinate system centered about the occlusal and midsagittal planes. Specifically, the subject wore a custom head-mounted jig during the experiment that contained 6 ireds whose positions were monitored continuously. In addition, 3 reference trials were collected prior to the experiment. For these trials, the subject held a plexiglass jig between his teeth.

Three ireds attached to the plexiglass allowed us to define a plane along the maxillary bite surface (occlusal plane).

The data were sampled at 150 Hz. and an acoustic recording of the voice was simultaneously digitized to serve as a reference during segmentation of the movement signals. The data were processed after the experiment to transform the lip data to a head coordinate system (Horn, 1987). The origin of the new head-coordinate system was the intersection of the midsagittal plane, the occlusal plane and an orthogonal plane running throughout the maxillary incisor cusp. The reference trials allowed us to define the origins of the coordinate system and the relation of the ireds on the head-jig to this system.

The lip data, with the head component removed, were examined using a waveform editor to determine the onset and end of the movements for the monosyllable. A crude estimate of oral aperture was computed by subtracting the vertical movement component of the midsagittal ired on the lower lip from the vertical movement component of the midsagittal ired on the upper lip. This signal was smoothed using a software-implemented Butterworth filter with a 15 Hz. cutoff frequency. The signal was then differentiated using a central difference algorithm and the zero crossings at the beginning and end of the syllables were identified.

Each syllable consisted of 24 movement streams for each trial (8 ireds times 3 spatial dimensions). The movement streams varied in duration from record to record with the number of sampled points per record ranging from the low 30's to a high of 51 (approximately 207 to 340 msec). To simplify data analysis, the data were interpolated so that each record had 51 equally spaced observations, and the time values 0, .02, .04, ..., 1 were assigned. No signal processing was carried out on the signal streams prior to the functional data analyses.

3 Statistical Methods

In this section three data decomposition or data partitioning methods are developed. The first, spline smoothing, permits a separation of variation at the ends of the defined interval from variation within the main part of interval. The second procedure, a functional version of one-way analysis of variance, permits the study of differences between syllables. The third, principal components analysis, within and across marker locations in terms of its dominant or principal features.

3.1 Notation

The subscript i will be used to indicate a record or replication ($i = 1, \ldots, 20$), subscript j will indicate a particular ired position ($j = 1, \ldots, 8$), and k specifies a particular time value ($k = 1, \ldots, 51$). Where necessary, a superscript v will indicate a specific vowel among the possibilities "bab", "beeb", "boob" and "bob" ($v = 1, \ldots, 4$). The notation t_k refers to time values, and $X_{ijk}^{(v)}, Y_{ijk}^{(v)}, Z_{ijk}^{(v)}$ will be used to refer to the values of the X-, Y-, and Z-coordinates, respectively, for record i, ired j, time value k and vowel v. Continual use of all these indices would imply rather unattractive and unreadable mathematical expressions, so any subscript not varying within a particular analysis will be dropped.

Some of the analyses involve the use of estimated derivatives of the coordinate functions.The notation Dx indicates the first derivative or velocity of coordinate function x, D^2x the second derivative or acceleration, and in general $D^m x$ indicates the derivative or order m. A specific value of, say, acceleration at time t is indicated by $(D^2x)(t)$.

3.2 Spline Smoothing and Decomposition

Although the noise level is small in these data, some degree of smoothing is essential to get good estimates of the first and second derivatives of the data. Smoothing serves another purpose in this paper: to partition or decompose each curve into two components, one measuring behavior at the end-points or near the boundaries of the curves, and the other describing their behavior in the central regions.

The basic idea behind spline smoothing is to define a function x that fits the observed data for coordinate X subject to a penalty placed on the lack of smoothness of x. The penalty function keeps function x from fitting the data precisely, but ensures that x has the appropriate amount of regularity or smoothness. In order to simplify notation, it is assumed that the data for a specific record, ired position, and vowel are being smoothed, and therefore the indices i, j and v are omitted in the following discussion.

The spline smoothing criterion for assessing the fit of smoothing function x_i for replication i of coordinate X is

$$Q_\lambda(X, x) = \sum_{k=1}^{51} [X_k - x(t_k)]^2 + \lambda \int_0^1 (D^4 x)^2(t)\, dt\ . \tag{1}$$

Of its two terms, the first term measures the badness of the fit of function

x evaluated at times t_k to the actual discrete data X_k in least squares terms: the closer the estimated function x passes to the data values X_k, the better the fit. In fact, if only this term were in the criterion, it would always be possible to find a function that fit the data exactly, and therefore reduced the criterion to zero. Such a function would be called an *interpolant* of the data.

The second measures the roughness of x, and its contribution to the criterion is to force x to sacrifice some fitting power in order to remain acceptably smooth. In this case roughness is measured in terms of the integrated or total squared fourth derivative D^4x. A function with limited variation in its fourth derivative will necessarily be smooth to some degree.

The amount of smoothness imposed by the second term is controlled by the penalty multiplier λ. The larger λ, the bigger the contribution of the second term, and therefore the more fit that must be sacrificed in order to keep the the term comparable in size to the first. It is instructive to consider the two limiting cases. As $\lambda \rightarrow \infty$ the size of the fourth derivative is ultimately forced to zero. This implies that the fitted function h would become a cubic polynomial, for which $D^4x = 0$ exactly. At the other extreme, as $\lambda \rightarrow 0$, less and less penalty is placed on smoothness, until finally the function x is able to fit the data exactly.

The actual smoothing parameter value used was $\lambda = 10^{-6}$, and was chosen by a process called generalized cross-validation (Wahba, 1990). That this value was so small is due to the raw data having very little noise variation, so that the estimated functions could very nearly fit the data exactly.

Smoothness is assessed in the second term in terms of the fourth derivative in (1) because we shall want to analyze the acceleration functions, D^2x, and the fourth derivative measures the curvature in the acceleration function. By controlling the net amount of curvature in acceleration one can ensure that the estimated acceleration is reasonably smooth.

3.3 End Point and Shape Variation

Although the criterion Q_λ above implies that the limiting fit for large λ is a cubic polynomial, it does not explicitly define what role this polynomial component would play for the penalty parameters of moderate or small size. In fact, we can choose this role explicitly, a feature that Ramsay and Dalzell (1991) suggested might contribute usefully to a functional data analysis. For the segmented speech movement data the movement variation between records for a particular ired tended to be of two kinds:

- end-point variation, or variation near the ends of curves, and
- shape variation, or variation in the central regions of the curves.

End-point variation is due in some degree to the fact that the utterance within which the syllable was embedded caused the lips to be positioned differently both at the beginning and ending of the syllable portion from record to record. Shape variation, on the other hand, is due to differences in the way the lips moved during the syllable, and is thus rather more important in this study. While these two types of variation cannot be considered to be entirely independent of each other, it can be useful to study them separately, in addition to studying the total curve.

The function x resulting from smoothing the data for a specific record, coordinate, ired and vowel (i, j, and v dropped) is split up as follows:

$$x(t) = u(t) + e(t) \tag{2}$$

where:

- u is the unique cubic polynomial for which values $u(0)$, $u(1)$, $(Du)(0)$ and $(Du)(1)$ match those of x at $t = 0$ and $t = 1$. This polynomial component captures end-point variation, but gives little information about changes within the interval because these four conditions use all of its degrees of freedom. Function u can be called the *endpoint* component of x.

- e is the function which has values and derivative values equal to 0 at the endpoints, but indicates the departure of the observed function x from polynomial u in the middle since $e = x - u$. Function e is therefore the *shape* component for a particular curve.

3.4 Principal Components Analyses

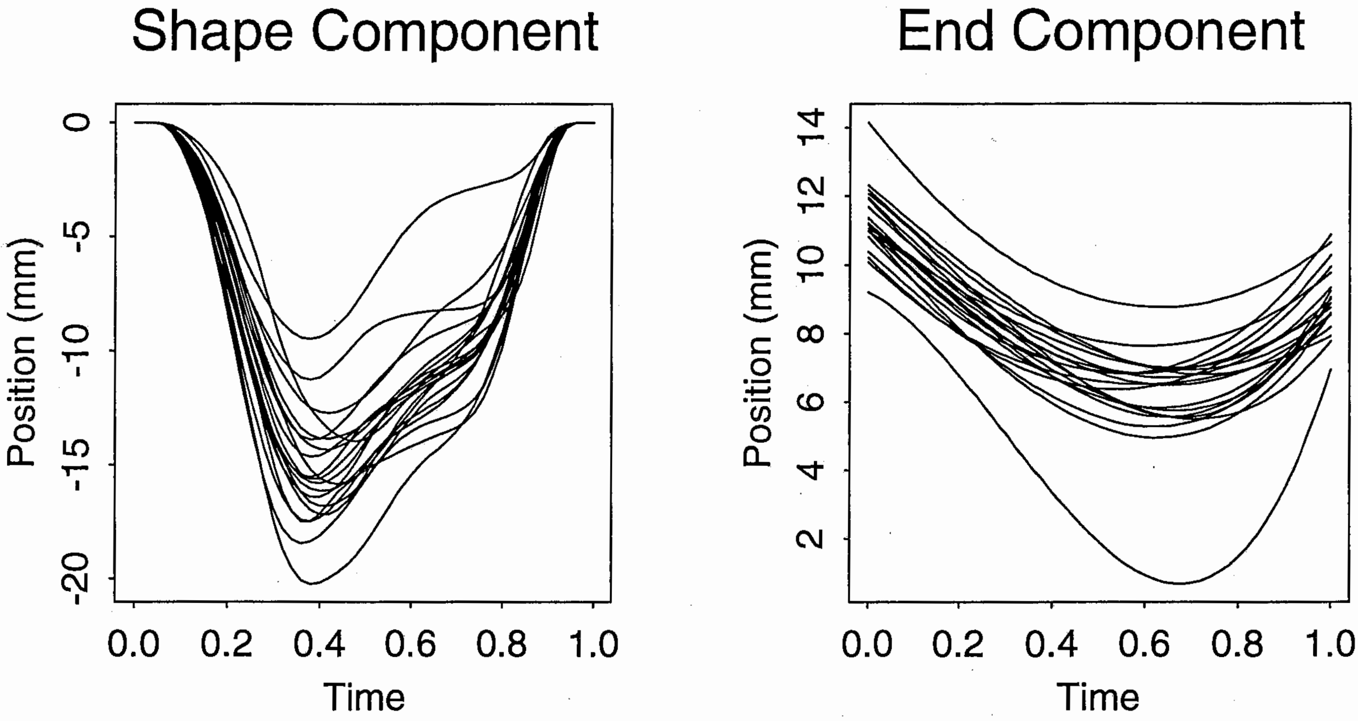
Principal components analysis (PCA) is used to explore the main modes of variation across records, and has many applications in this study. One of the most useful is to define a local coordinate system, the *principal axis* system, that can effectively replace the three spatial coordinates by one.

. Figure 1 shows the motion along the first principal axis for the lower-central ired for the syllable "bob" separated into its shape and end-effect components. With one exception, the end-effect u-components are very similar in shape, vary primarily in the starting and ending levels, and represent about 4 mm of movement per record. The shape *e*- components, on the other hand, have the essential features of the total records within the interval, but have fixed position and velocity at the ends.

Within a specific ired coordinate, one is interested in not only by how much the records vary, but also in the ways in which they vary. A critical question concerns how many important types or modes of variation the data display. This tends to indicate the complexity of the processes driving the system, such the neural processes controlling muscle response and the internal biomechanical constraints on tissue movement.

We can also use principal components analysis to explore variation across all three coordinates within a specific ired, and even the total simultaneous variation among the 24 ired coordinates. Again, a functional version of principal components analysis can be defined by fairly simple modifications of the multivariate version. If the multivariate data had values x_{ik}, with i indexes replications and k indexes variables, we would proceed, after first centering the data by subtracting the across-replication mean values $\bar{x}_k$ from each observation, to construct the cross-product or variance-covariance matrix $\mathbf{V}$

Figure 1: Twenty records of motion along the first principal axis for the lower-central ired split into shape effect e and end-effect u. The total curves are displayed in the left of Figure 6.



Shape Component

End Component

with entries

$$v_{jk} = N^{-1} \sum_i [x_{ij} - \bar{x}_j][x_{ik} - \bar{x}_k] \ . \tag{3}$$

The next step in a multivariate PCA would be to find the eigenvalues and eigenvectors of matrix $\mathbf{V}$, with the number of large eigenvalues indicating the number of dominant principal components, and the corresponding eigenvectors being used to give an indication of the nature of these components of variation. The eigenanalysis can be expressed as the problem of finding the solutions to the eigenequation

$$\mathbf{V}\mathbf{u} = \gamma \mathbf{u} \ , \ \mathbf{u}^t \mathbf{u} = 1$$

where γ is an eigenvalue and $\mathbf{u}$ is an eigenvector.

When an observation is a continuous function with values $x_i(t)$, index k being replaced by argument t as above (in this application actually taking on 51 discrete values t_k). In the multivariate context, one would usually also standardize by dividing each residual from the mean by the standard deviation, but standardization wouldn't make sense here, where all the "variables" have the same scale. The covariance matrix $\mathbf{V}$ in the multivariate context now becomes, in the functional context, the bivariate function

$$v(s,t) = N^{-1} \sum_i [x_i(s) - \bar{x}(s)][x_i(t) - \bar{x}(t)] \ , \tag{4}$$

where

$$\bar{x}(t) = N^{-1} \sum_i x_i(t) \ .$$

The eigenenalysis problem now has the form

$$\int v(s,t)u(t)dt = du(s) \ , \ \int u^2(t)dt = 1 \ ,$$

where d is still called an eigenvalue of v, but where function u is the corresponding *eigenfunction.* It is apparent that summation over index k has now been replaced by integration over continuous index t.

Actually, however, if the functions are observed at equally-spaced discrete values t_k, with $\delta = t_{k+1} - t_k$, then the values for all replications can be tabled in a data matrix $\mathbf{X}$ having 20 rows, each corresponding to a record, and 51 columns.

The integrations can be well approximated by the trapezoidal rule for numerical integration, having the form

$$\int v(s,t)u(t)dt \approx \delta[\sum_k^{51} v(s,t_k)u(t_k) - (v(s,t_1)u(t_1) + v(s,t_{51})u(t_{51}))/2]$$

and

$$\int u^2(t)dt \approx \delta[\sum_k^{51} u^2(t_k) - (u^2(t_1) + u^2(t_{51}))/2] \ .$$

Except for multiplication by the constant δ and the small correction resulting from subtracting the average of the end-point values, this puts us back to the matrix computation (3). The impact of the end-point correction in this rule is rather small for as many as 51 time-values, and the multiplication by the constant δ doesn't change the interpretation of results, so it is fair to say that, with a reasonable number of equally spaced time values, doing a conventional multivariate PCA of the discrete function values is almost correct. The results reported here, however, do make use of the trapezoidal rule.

One also wants to conduct an analysis across both coordinates and ireds in order to investigate the synergies or coherences between motions of the eight lip positions. The vector function

$$(x_{i1}, x_{i2}, \ldots, x_{i8})^t$$

has eight components and the corresponding data matrix $\mathbf{X}$ has $8 \times 51 = 408$ columns.

In all three analyses, the number of replications, 20, is much less than the number of columns of the data matrix $\mathbf{X}$, and therefore the number of nonzero eigenvalues that can be computed is no more than 20.

3.5 Functional Analysis of Variance

We shall need to explore the systematic differences among lip position functions $x^{(v)}$, as well as their acceleration counterparts, across the four experimental syllables ($v = 1, \ldots, 4$). Ramsay and Dalzell (1991) discuss the functional linear model in general, and functional analysis of variance in particular, although in a context rather more general than needed here. The following discussion is simplified by the fact that there are equal numbers of observations for the four syllables.

If the syllable comparison problem were the classic one of studying the across-treatment variation of a simple one-dimensional variable y with the value y_{ij} for replicate i within treatment j, then the one-way analysis of variance (ANOVA) model would be

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

In this model parameter μ is the grand mean across treatments, α_j measures the unique contribution of treatment j, and e_{ij} is a residual or error term. The constraint

$$\sum_j \alpha_j = 0$$

is usually imposed to ensure that the treatment effects are uniquely defined.

If the observation were multivariate in character, with values y_{ijk}, with indices i and j as above, but with the added index k indexing variables, the ANOVA model extends to multivariate or MANOVA model

$$y_{ijk} = \mu_k + \alpha_{jk} + e_{ijk} \ .$$

Here, however, we are interested in across-syllable variation of the position functions with values $x_i^{(v)}(t), y_i^{(v)}(t)$, and $z_i^{(v)}(t)$ and their derivatives. This implies the counterpart functional ANOVA, or FANOVA model

$$x_i^{(v)}(t) = \mu(t) + \alpha^{(v)}(t) + e_i^{(v)}(t) \tag{5}$$

in which the continuous variable t has replaced the discrete index k and the treatment subscript j has switched to syllable superscript v. Function μ represents the grand mean position for all records and treatments, and the functions α_v specify what is unique in position variation for specific syllables v. The corresponding identifiability constraint is

$$\sum_v \alpha_v(t) = 0 \text{ for all } t \ . \tag{6}$$

It turns out that most of the computational procedures and goodness of fit summary statistics used in ANOVA can be transported with relatively obvious changes to accommodate this functional context. To estimate the across-syllable mean μ and within-syllable effects α_v one proceeds as follows. Making use of the fact that the sample size $N = 20$ is the same for each syllable, the parameter estimates are

$$\begin{aligned}\hat{\mu}(t) &= (4N)^{-1}\sum_i\sum_v x_i^{(v)}(t) \\ \hat{\alpha}^{(v)}(t) &= (N)^{-1}\sum_i x_i^{(v)}(t) - \hat{\mu}(t) \ .\end{aligned} \tag{7}$$

Residual functions $\hat{e}_i^{(v)}$ are then estimated by

$$\hat{e}_i^{(v)}(t) = x_i^{(v)}(t) - \hat{\mu}(t) - \hat{\alpha}^{(v)}(t) \ .$$

¿From the residual functions one defines the error sum of squares functions

$$SSE(t) = \sum_i\sum_v[\hat{e}_i^{(v)}(t)]^2 \ .$$

Two useful summary functions are the squared correlation function

$$R^2(t) = [SSE_0(t) - SSE(t)]/SSE_0(t)$$

and the F-ratio function

$$F(t) = \frac{[SSE_0(t) - SSE(t)]/3}{SSE(t)/76}$$

where SSE_0 is the null hypothesis error sum of squares,

$$SSE_0(t) = \sum_i\sum_v[x_i^{(v)}(t) - \hat{\mu}(t)]^2 \ .$$

For a fixed value of t, $F(t)$ has, in this application, numerator and denominator degrees of freedom 3 and 75, respectively.

4 Displays and Analyses of Lip Position

4.1 Descriptive Displays and Analyses

Figure 2 shows the 3 coordinate functions for the ireds positioned at the upper-center, upper-right, extreme-right, and lower-right, respectively, and Figure 3 shows the records for ireds positioned at the lower-center, lower-left, extreme-left, and upper left, respectively. The X-direction is vertical position, the Y-direction is lateral position, and the Z-direction is protrusion or fore/aft position. It is apparent from these plots that most of the movement is in the X- and Z-coordinates for the lower lip ireds. The lower central ired, for example, typically moves about 25 mm vertically, 4 mm fore and aft, and only 1 mm laterally. It should be appreciated that a large part of this movement is contributed by jaw motion; our data did not permit a separation of relative lip position from jaw position. We will separate lip from jaw motion in subsequent papers.

Figure 2: Twenty records for X, Y, and Z coordinates for ired positions at the upper-center, upper-right, extreme-right, and lower-right, respectively, during the utterance of the syllable "bob". The X-direction is vertical position, the Y-direction is lateral position, and the Z-direction is protrusion or fore/aft position, all in millimeters.

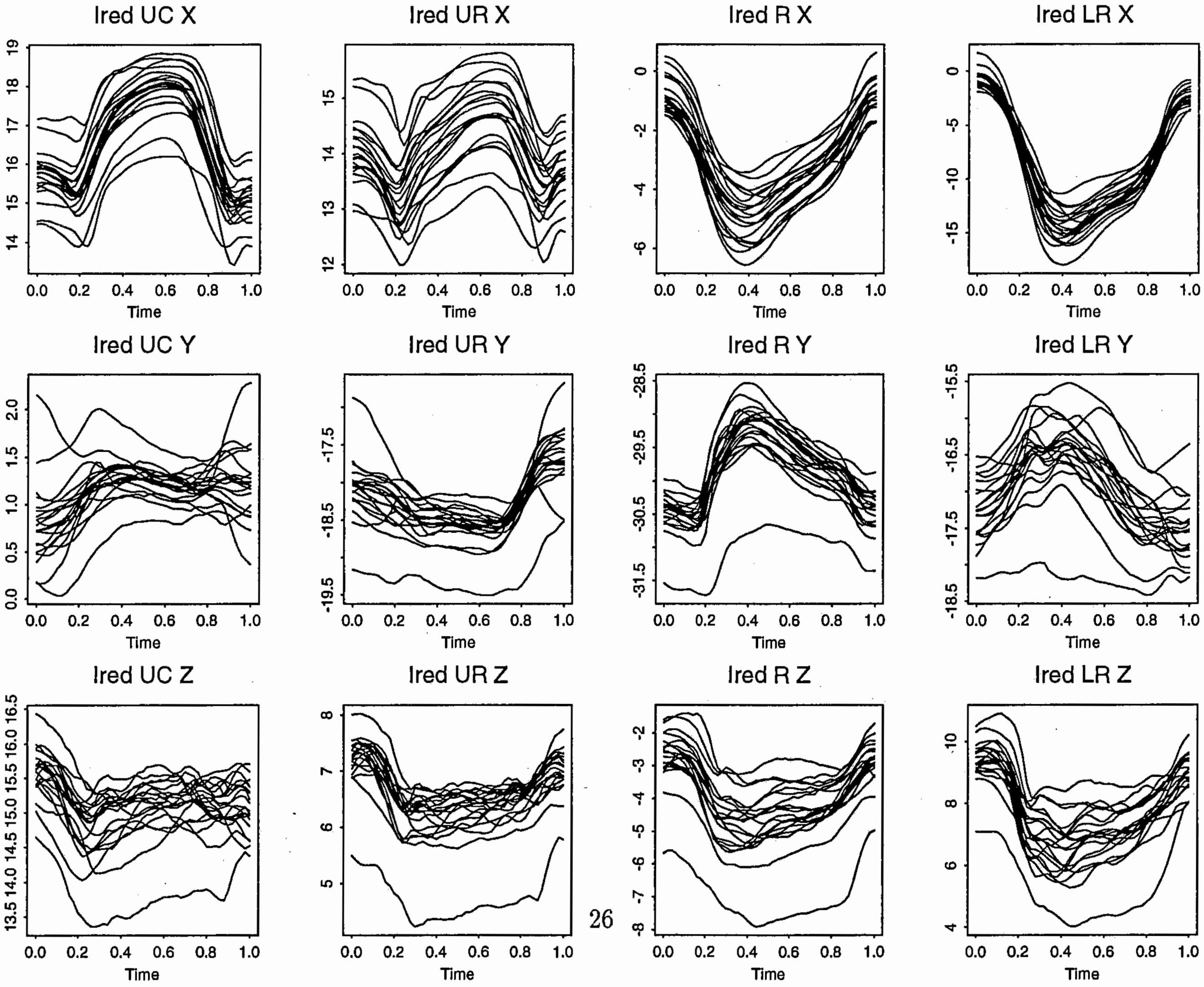


Figure 3: Twenty records for X, Y, and Z coordinates for positions are lower-center, lower-left, extreme-left, and upper left, respectively, during the utterance of the syllable "bob".

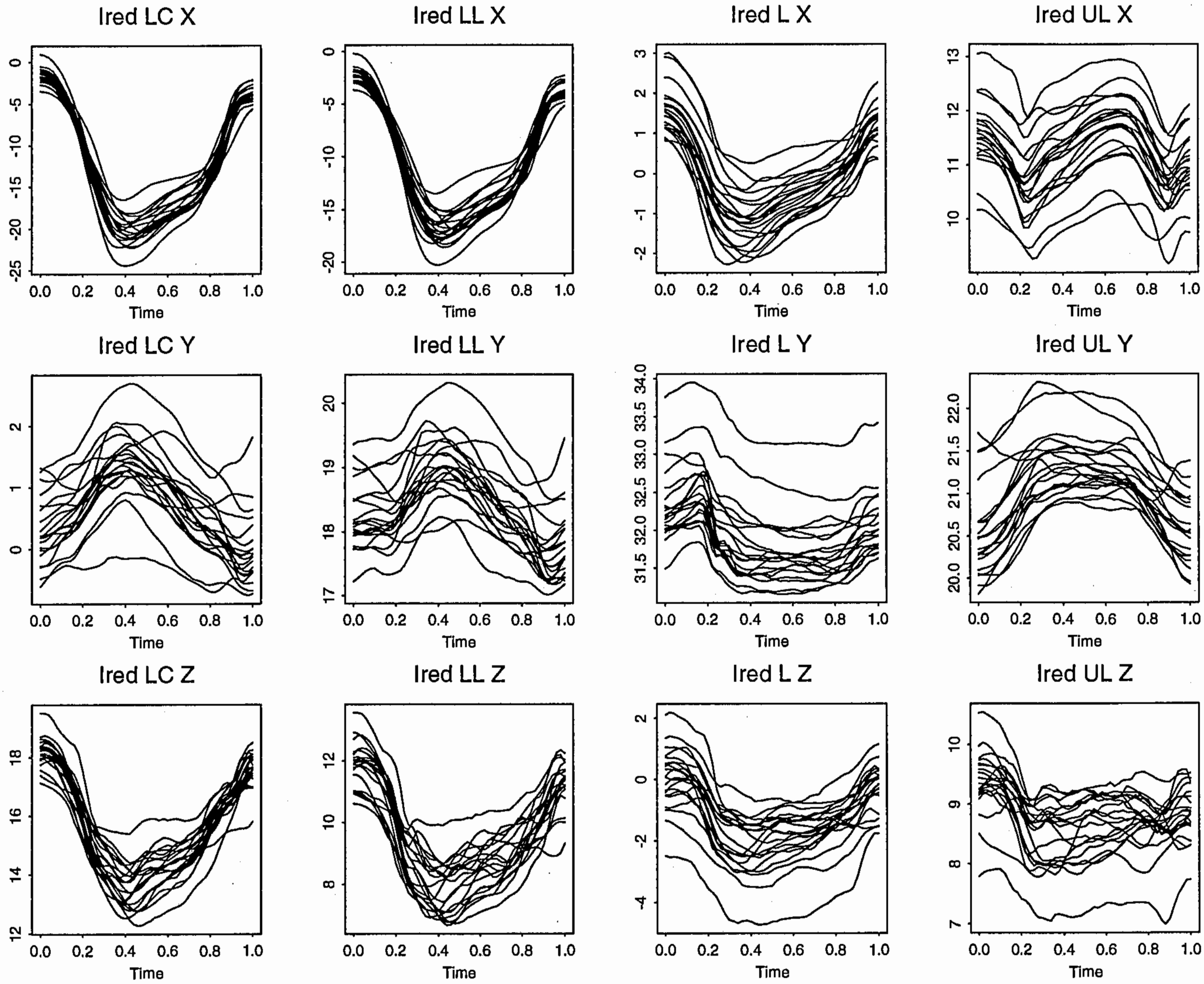


Figure 4 plots the mean ired positions in the sagittal (X versus Z) and frontal (Y versus X) planes. These plots show that the movement is mainly in the up-down X-direction, with some in the fore-aft Z-direction and little in the lateral Y-direction. The lower lip, which begins by protruding a little beyond the upper lip, flips down and back fairly sharply, followed by a slower return to nearly the original position. What little that happens laterally takes the form of a slight movement of the corners of the mouth inward as a concomitant to the sagittal motion.

Figure 5 displays for comparative purposes the mean movement of all 8 ireds for each syllable in the sagittal plane. The syllables "bab" and "bob" involve greater movement of the lower lip than "beeb" and "boob", and "boob" involves more forward or protruding lower lip movement than the others.

4.2 Principal Axis Transformation

The fact that within-ired motion is nearly linear suggests that the 3D motion of single-ired features can be well represented in the line or plane defined by the first one or two principal components of variation, respectively. This

Figure 4: The left figure displays sagittal plane lip motion for mean lip positions for all 8 ireds during "bob", and the right figure displays frontal plane movement. In the sagittal plane display, the dashed straight lines centered on the lower-central curve indicates the principal directions of variation for this ired, or its local principal coordinate system.

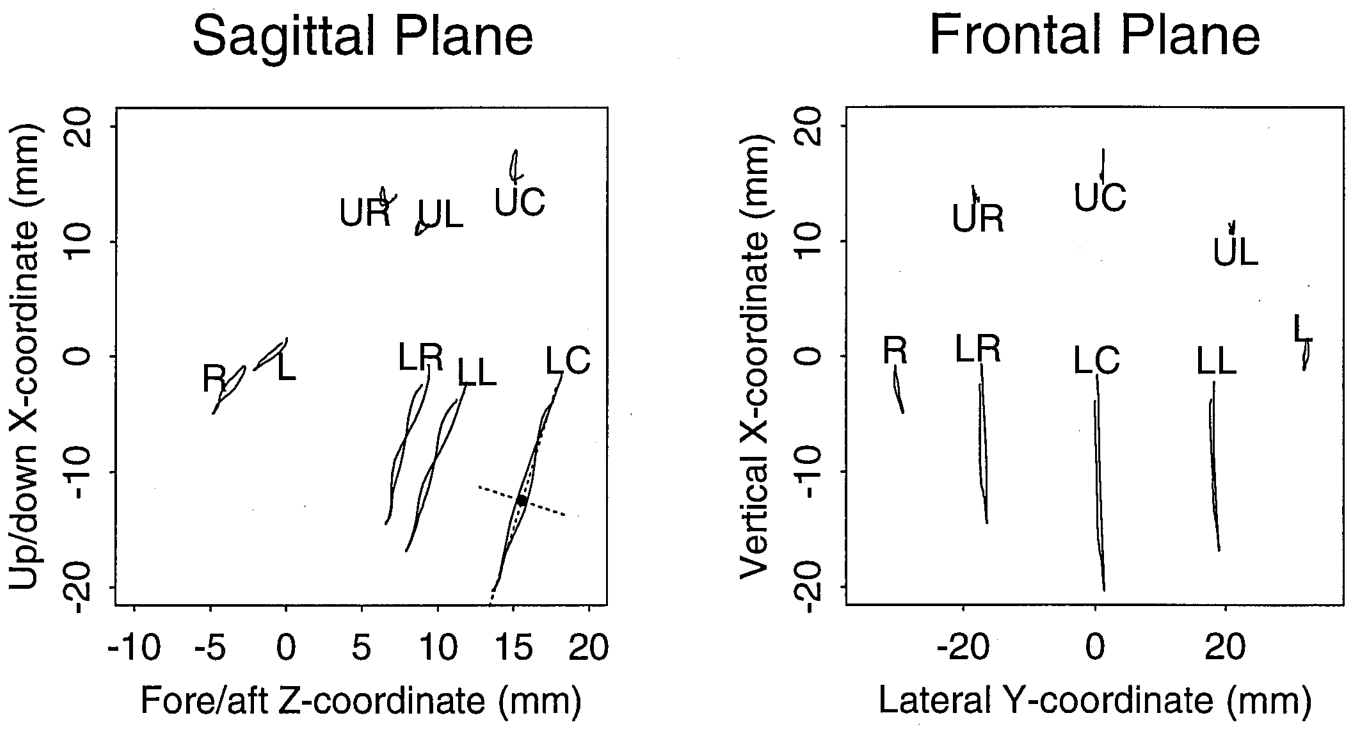
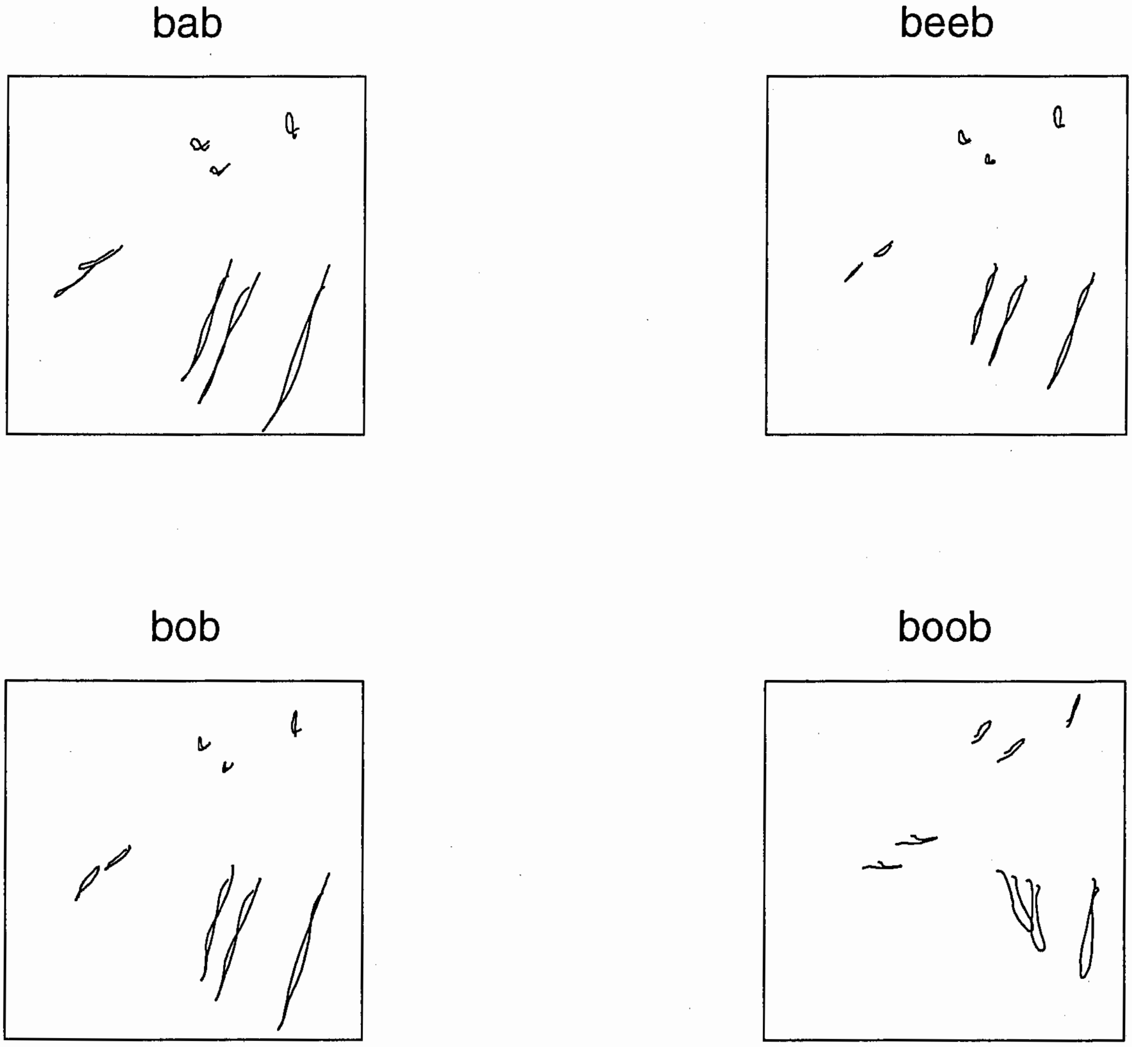


Figure 5: Sagittal plane mean motion for each syllable and all ireds.



bab

beeb

bob

boob

variation is taken with respect to the mean coordinates defined by averaging within an ired across time. These principal components define a best local coordinate system for displaying that ired's effects, and will be called its *principal axes.* These principal axes are obtained for a specific ired by computing the eigenvectors of the order 3 variance-covariance matrix for ired coordinates, and transforming the 51 times 3 matrix of centered ired coordinates by the matrix formed by using the first one or two eigenvectors.

For example, the eigenvalues of the variance-covariance matrix for the mean lower-central ired are 44.25, 0.08, and 0.02, implying that motion in the first principal component direction accounts for 99.78% of the variation, and that the least important direction accounts for only 0.05% of the motion, and thus can be ignored for plotting purposes. A display of the three coordinates $(x(t), y(t), z(t))$ with respect to the first two axes of this local coordinate system for this ired is achieved by first subtracting the ired centroid vector $(-12.47, 0.71, 15.54)^t$ from these functions, and then multiplying by the partial rotation matrix

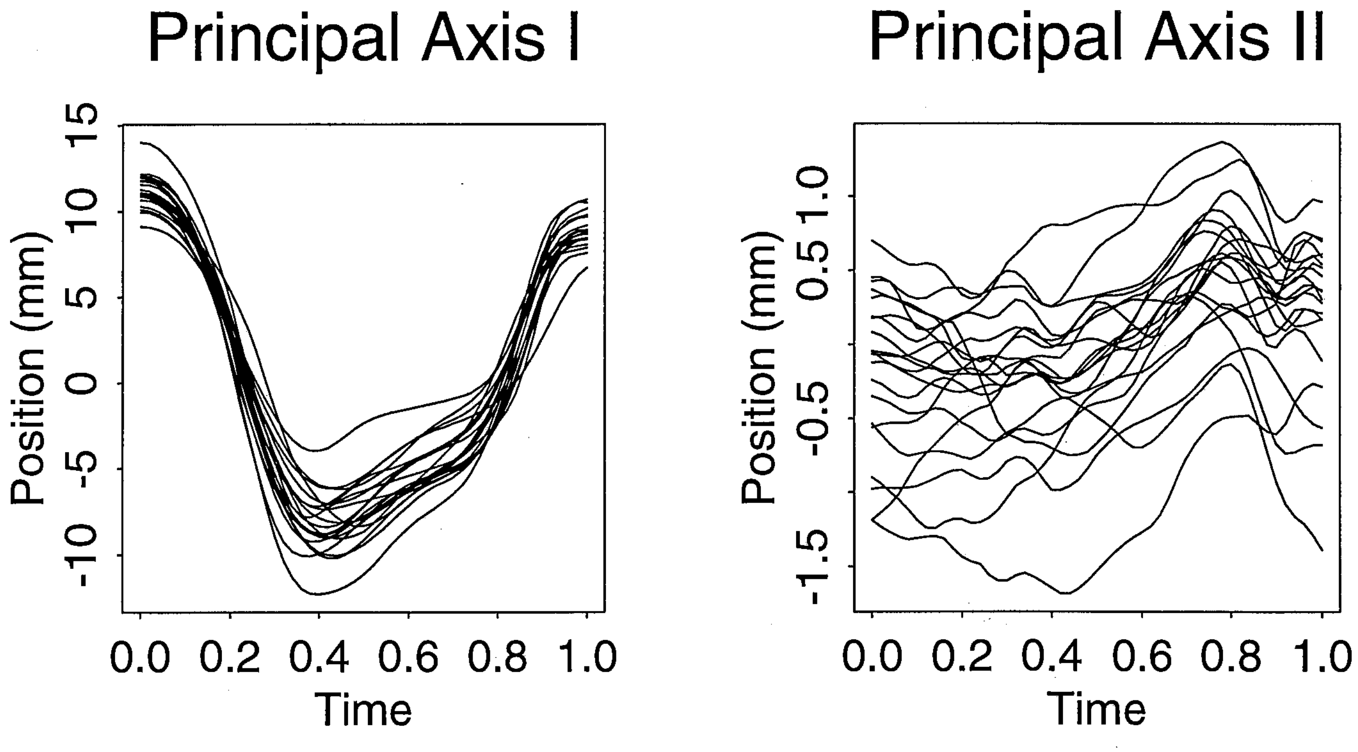
$$\begin{bmatrix} 0.97 & -0.12 \\ -0.06 & -0.95 \\ 0.22 & 0.28 \end{bmatrix}$$

The axes of this local coordinate system are displayed for the lower-central ired in Figure 4. The motion of this ired along the first two principal axes is displayed in Figure 6. One notes in the right plot that the movement along the second axis, accounting for only 0.2% of the motion, is relatively chaotic, further suggesting that, for the purposes of this paper at least, it may be ignored. Along the dominant axis the motion of the lower central ired passes through essentially three phases: a first phase lasting until $t = 0.3$ in which the lip drops rapidly, as second phase until about $t = 0.7$ in which the lip is closing slowly and rather linearly, and a third concluding phase of more rapid closure.

4.3 Some Observations on Lip Motion Control

When a time-varying process is under some kind of external control that is distributed over an extended interval of time, there is usually a strong tendency for the covariance function to become large, either positively or

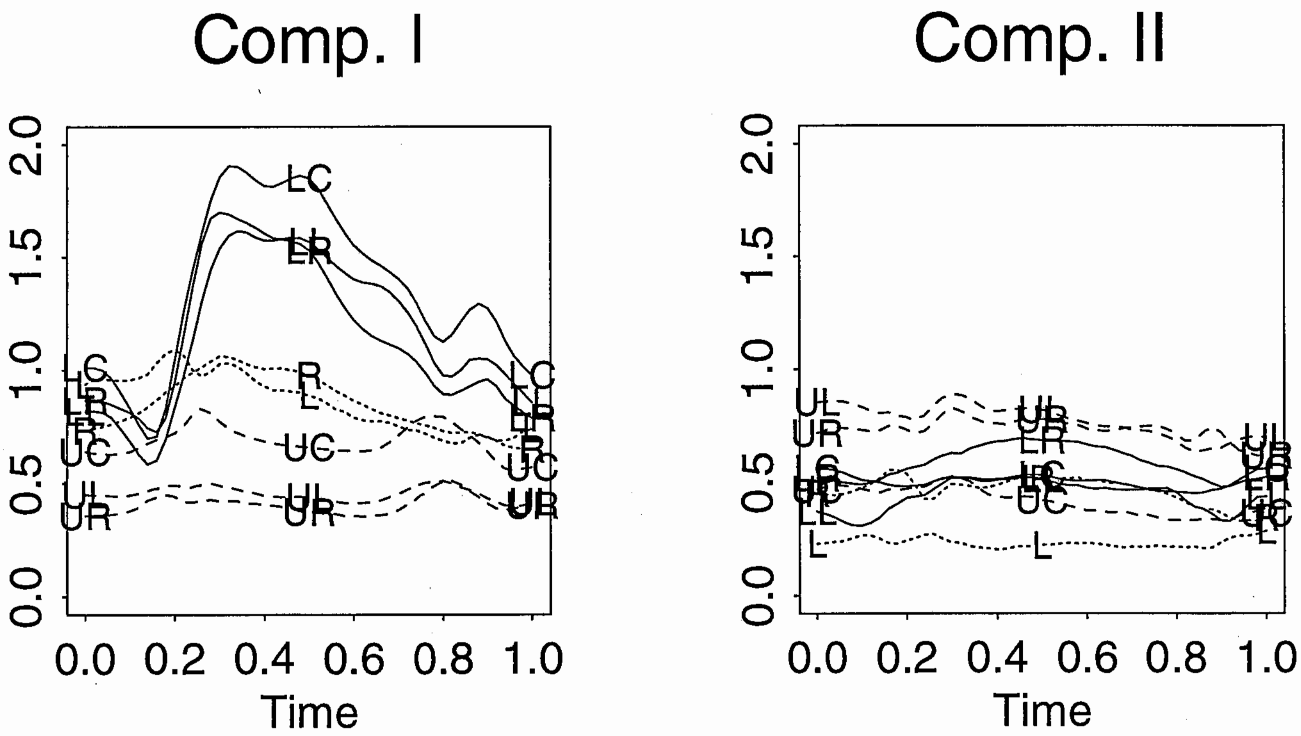
Figure 6: The left plot shows the motions of the lower central ired along the principal and dominant axis of motion for "bob" relative to the mean position. The right plot displays the motion along the sub-dominant second principal axis of motion.



negatively, and the effects of the control spread along the time-axis. By contrast, when the system is not being driven from outside, its behavior is determined by its internal dynamics, which are as a rule short-termed in their effects. For example, passive tissue tends to have strongly spring-like characteristics, and its behavior is due largely to its spring-constant. One way to examine the different influences on lip motion is to evaluate, in a functional sense, the standard deviations and the correlations within and among the ireds. The standard deviation curves in Figure 7 plot the standard deviation of the position of each ired as a function of time along its two principal axes of motion. They indicate that the ired coordinates with the largest motion also have the largest standard deviation across records, and that the standard deviation is also greatest along the first principal axis or direction of motion. There is a background or baseline standard deviation of around 0.5 mm in all records.

The correlations among ired positions for different values of time define a set of bivariate functions of time. Let $r(t_{k_1}, t_{k_2})$ denote the correlation between ired positions at times t_{k_1} and t_{k_2} for a specific position function. The resulting matrix of correlations is of order 51 for these data, and therefore impractical to display. But since the correlation will vary smoothly as a

Figure 7: Standard Deviation of lip positions for all 8 ireds during "bob". The left display is for motion along each ired's first and dominant principal axis, and the right display is for the second principal axis. Curves for the lower right, central and left ireds are solid lines, for the upper right, central and left are dashed, and for the extreme left and right are dotted.



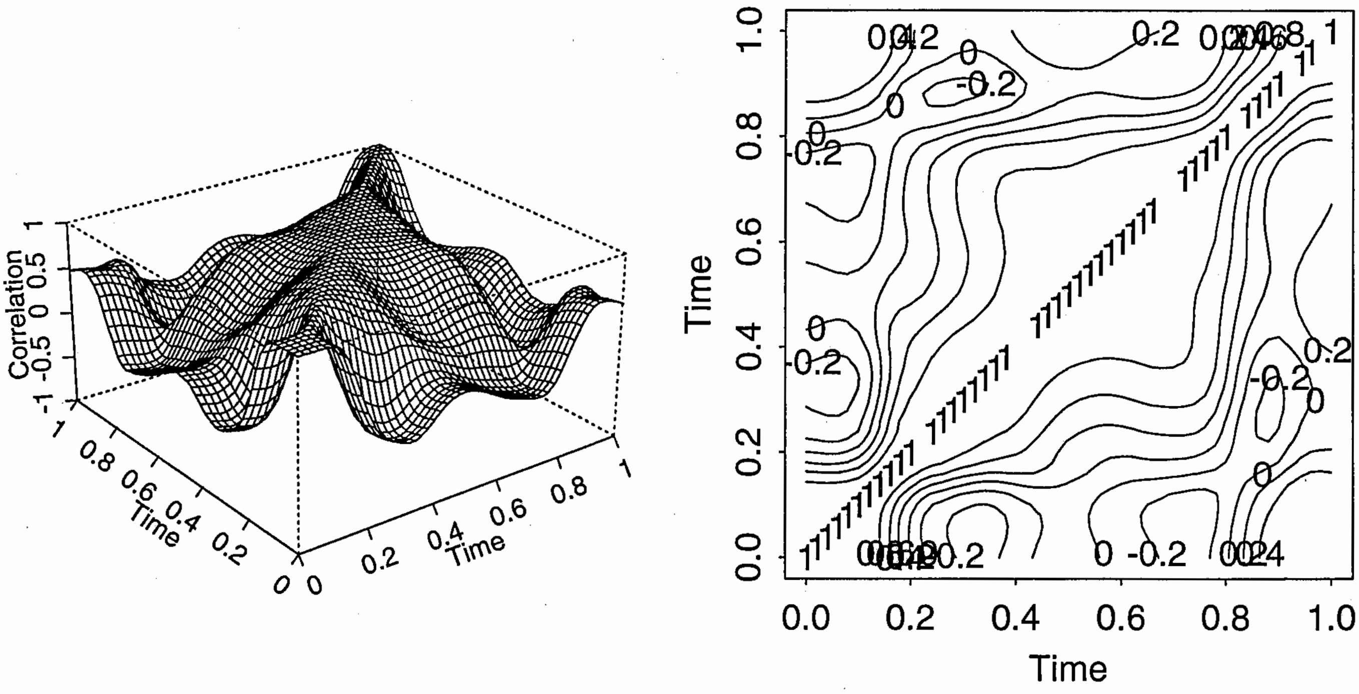
function of the two time values, these correlation values can be displayed as a surface over the time by time plane.

Figure 8 shows the correlation surfaces for movement along the first principal component of movement for the lower-central ired as a perspective plot and as a contour plot of the surface. In the perspective plot we see the diagonal ridge running from foreground to background that contains the unit correlations for equal time values, and this ridge shows up as running from lower left to upper right in the contour plots.

Of particular interest is the manner in which correlations fall off on either side of the diagonal ridge as one moves from the beginning to the end of the time interval. In notational terms, this means looking at the correlations $r(t+\delta, t-\delta)$: the value of t gives the position along the diagonal ridge, and the value of displacement δ gives the distance from the top of the ridge along a line perpendicular to it.

Near the ends of the interval the correlations fall towards zero rather rapidly, as evidenced by the dense contour lines. But there is a flat spot at about the two/thirds point ($t = 0.6$) where correlations stay high for fairly widely separated values. To understand this effect, it is necessary to take

Figure 8: The correlations among lip positions along the dominant principal component of movement for the lower center ired are displayed as a perspective surface plot on the right. The diagonal ridge running from front to back contains the unit or near unit correlations for pairs of very similar time values. Note the flat region just after the middle of the interval. The left plot displays the correlation surface in a contour plot.



into account both the standard deviation $\sigma(t)$ and covariance $\sigma(s,t)$ since

$$r(s,t) = \sigma(s,t)/[\sigma(s)\sigma(t)] \, .$$

A comparison of Figure 8 with Figure 7 indicates that covariance in this region is elevated relative to the standard deviation. These flat regions in the correlation surface, then, seem to indicate that the system is under external or exogenous control, and these flat regions may thus correspond to the time of activation of the muscles associated with lip and jaw closure. Similar effects were noted in Ramsay (1982).

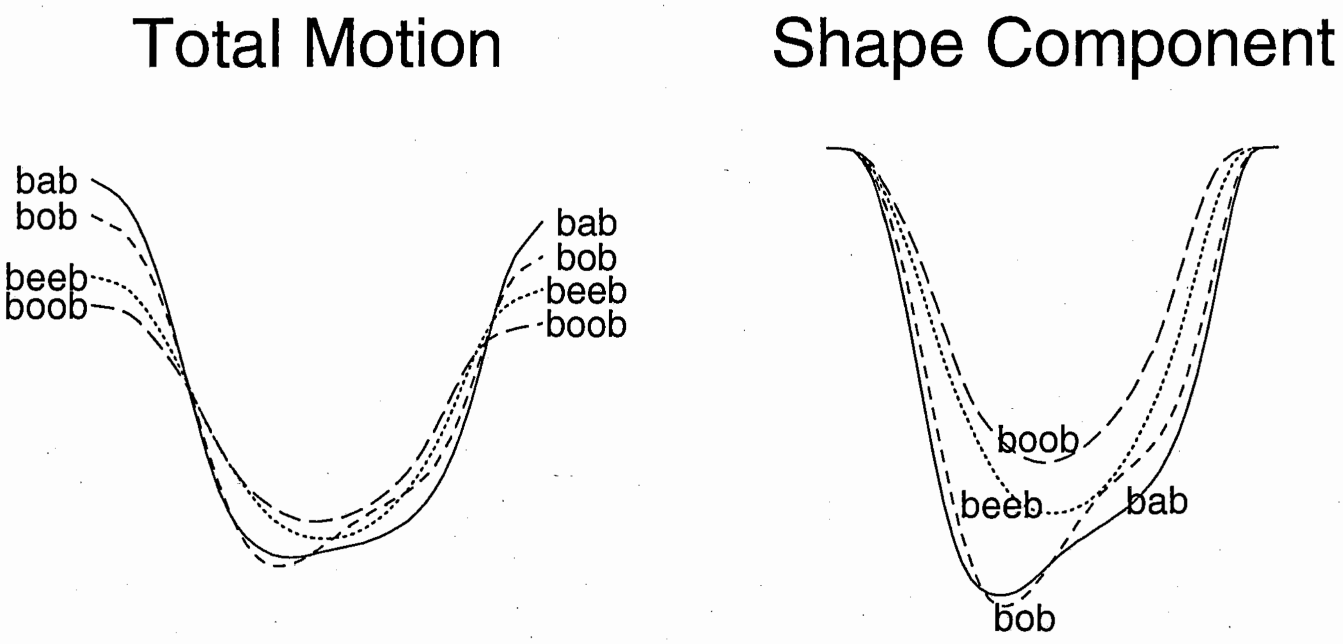
4.4 Functional Analysis of Variance Results

It is clear from the sagittal plane plot in Figure 5 there there are important differences in the average motion of the ireds across syllables in terms of the principal axes of motion. This plot does not permit us to see, however motion along these axes tends to differ systematically from syllable to syllable. The left of Figure 9 displays the mean trajectories for the four syllables along the principal axis of motion specific to each syllable for the lower central ired. There would appear to be important differences, so that, for example, the amount of motion is rather larger for “bab” than for “boob” for this ired.

One notes two nodes where all four trajectories tend to coincide. The right part of Figure 9 displays only the shape effect, and gives a better idea of how the trajectories differ once end position differences are removed. Syllables “beeb” and ”boob” not only show less motion than the other two, but also exhibit less asymmetry in their trajectories.

In order to confirm the differences noted in Figure 9 are substantial in a statistical sense, a functional analysis of variance of the motion of each ired along that ired’s principal axis of motion was carried out, for the total motion and for the shape component. The strength of the inter-syllable variation is summarized in Figure 10 in terms of the squared correlation $R^2(t)$ as a function of time for each ired. The left display shows the effects for total variation, while the right display shows the effects for the shape components. The value of R^2 needed to achieve significance at the 5% level is 0.10, and is indicated in the Figure. We can see that the amount of inter-syllable variation in total motion is large at the ends and in the middle of the interval, but falls close to insignificance at the two points of sharp acceleration, $t = 0.3$ and $t = 0.8$. The lower central ired stands out as having limited inter-syllable variation in the center of the interval as well. The shape components, however, have substantial variation over the rest of the interval,

Figure 9: The average position of the lower central ired along the principal axis of motion is plotted for each syllable in the left display. In the right display only the shape component is plotted.



Total Motion

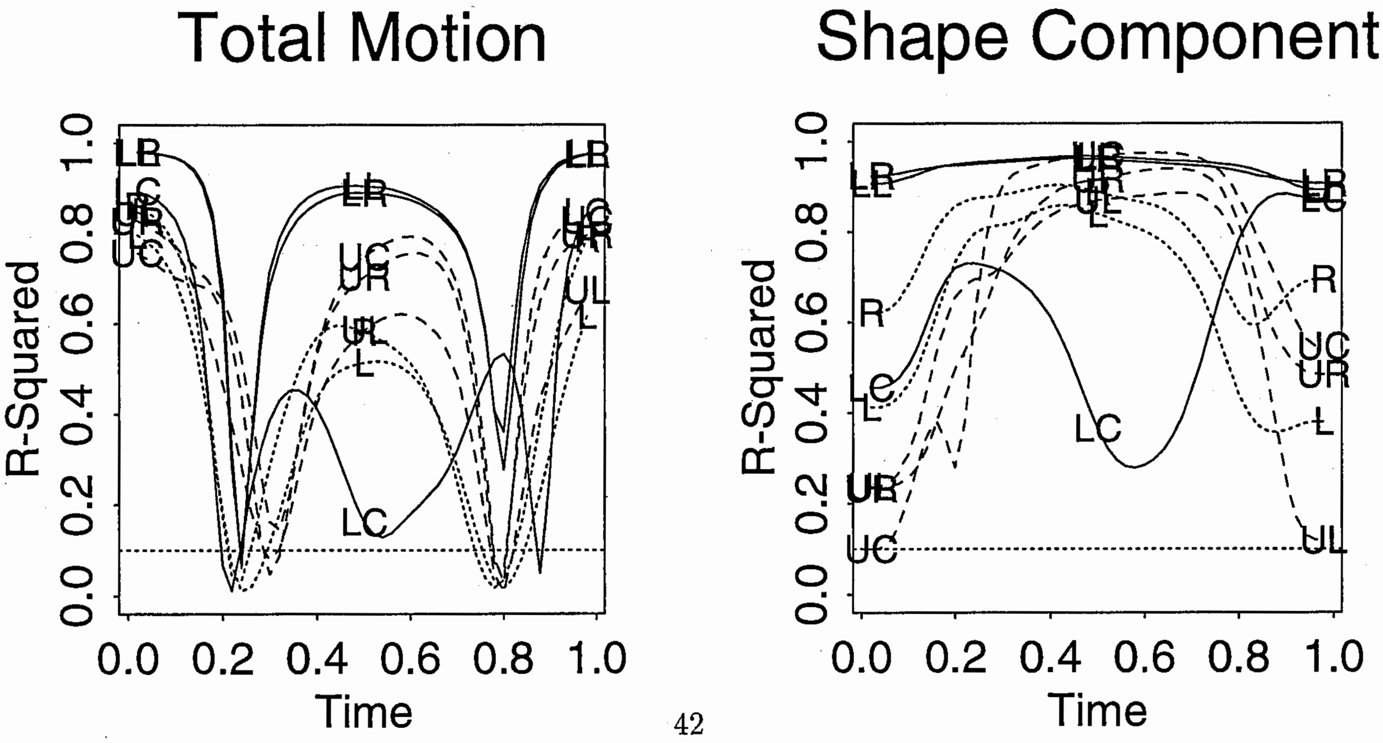
Shape Component

including at the two acceleration episodes.

4.5 Principal Components Analyses

A central question in this analysis concerns the dimensionality exhibited by the motion of the eight lip positions, each involving three coordinates. There would be in principle the potential for complex and high-dimensional variation in individual trajectories, both within an ired for its three coordinate functions, and also between ireds among the 24 coordinates as a syllable was articulated. As a first step, principal components analysis was undertaken to reveal the complexity of variation for each ired independently along its principal axis of variation and about the mean curve. A PCA was carried out separately for the total variation and for the shape variation. Table 1 gives the proportions of variances accounted for by the first three principal components for "bob". The results of the analysis for each ired separately indicated that the variation around the mean trajectory over replications was strongly one-dimensional for this syllable, especially for the shape component for the lower lip ireds, as seen in Table 1, where motion along the single dominant trajectory accounted for about 87% of the variance. The primary

Figure 10: The left display of shows the squared multiple correlation R^2 as a function of time for the analysis of between-syllable variance for all ireds and for total motion. The right display is for shape component only. Curves for the lower right, central and left ireds are solid lines, for the upper right, central and left are dashed, and for the extreme left and right are dotted. The 5% significance level for R^2 for 3 and 75 degrees of freedom is shown as a dotted horizontal line.



type of variation was simply how wide the lips were opened, with the shape of the trajectory remaining relatively unchanged. This stability of the shape component over replications is visible in Figure 1 for the lower-central ired.

The first two columns indicate that the first component strongly dominates all others for both total curve variation and shape variation. The dominance of the first component is stronger for shape variation than total curve variation for the lower lip ireds, suggesting that some of the lower lip variation is due to variation at the ends of the intervals. For the upper and left lip ireds, however, there is a tendency to see more than one dimension of shape variation. Nevertheless, the first principal component is strongly dominant for all ireds.

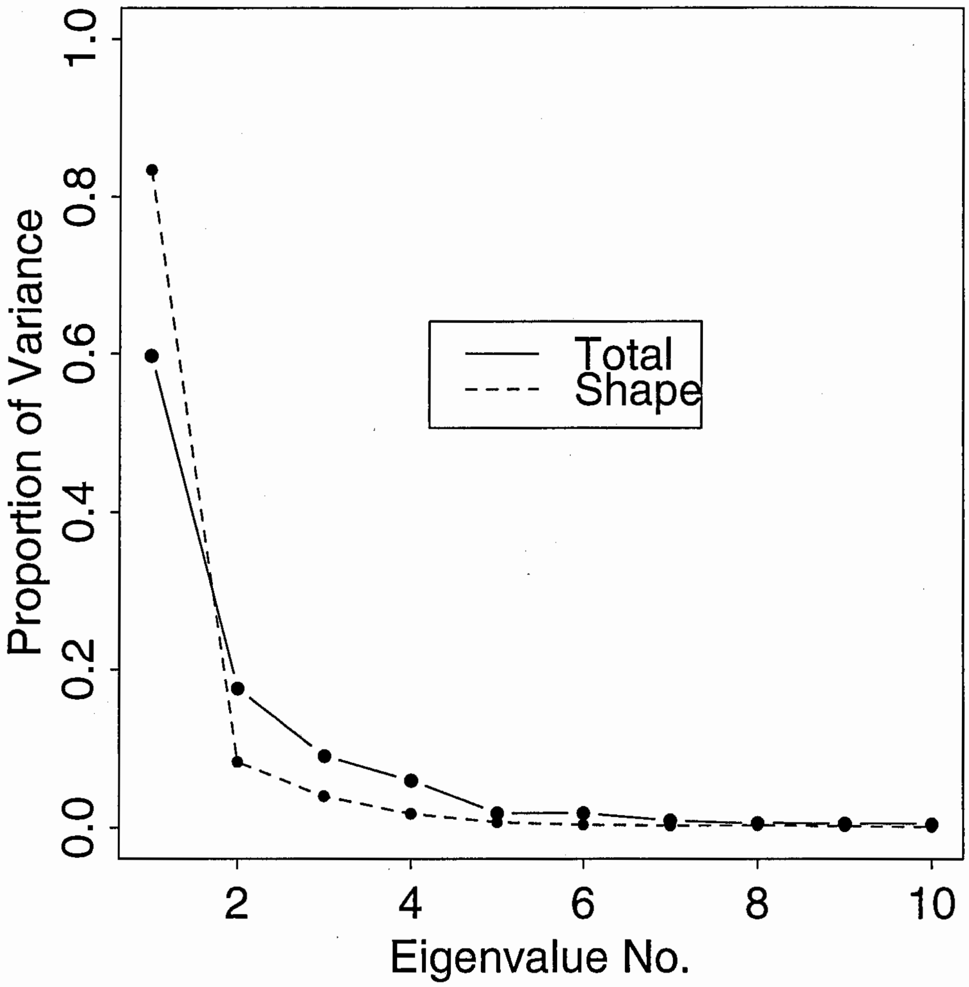
The principal components analysis of combined variation of the ireds along their respective first principal axes of motion reveals the complexity of simultaneous variation. It is entirely possible for two or more ireds to separately have only one component of variation, but for the simultaneous variation to be more complex, just as a single variable by definition has only one component of variation, a collection of variables can exhibit quite complicated and high-dimensional patterns of combinbed variation. The proportions of variance accounted for by the across-ired principal components are displayed in Figure 11 for both total curve variation and for shape variation alone. The principal components analysis of simultaneous ired motion, especially when only shape components were used, suggested strongly that most of the variation in lip motion across ireds was contained within a single dominant dimension of variation. This is perhaps not too surprising, given that most of the motion is in the three lower lip ireds, and it was already observed that each of them moved in an essentially one-dimensional trajectory.

Principal components results for the other syllables analyzed separately were essentially the same in terms of the features found for "bob".

5 Displays and Analyses of Lip Acceleration

It is profitable to look closely at the acceleration of the system because acceleration is proportional to force. The spline smoothing procedure permitted us to estimate the acceleration of the lip ireds during articulation, and it seems likely that the various controlling processes determining lip motion will have their most visible impact on acceleration. Hence, it is in the study of acceleration rather than position that one probably come closest to studying the processes controlling lip motion. These analyses did not separate

Figure 11: The proportions of variance accounted for by the first ten principal components of simultaneous or joint ired variation along each ired's first principal axis for "bob".



shape from end-effect variation.

Figure 12 shows the average acceleration functions for each of the ireds along their respective principal axes of variation. The lower-central X-coordinates (solid lines) have strong positive accelerations at around $t = 0.3$ and $t = 0.8$. The lip is first accelerated downward, then slowed down, followed by a short period ($t = .5$) of near zero acceleration, followed by a strong acceleration upward and then finally a negative acceleration as the lip returns to the closed position.

The other ireds show less acceleration and more complex patterns.

Figure 13 indicates the variation in the lower central ired acceleration across syllables in their respective first principal axes of motion. The longer syllables, "beeb" and "boob", fail to exhibit the momentary period of zero acceleration of the shorter syllables, and exhibit less acceleration or smoother motion throughout their trajectories.

A functional analysis of variance of the across-syllable acceleration variation was carried out for each ired, in the same manner as was done for position. The squared correlation functions for the lower lip ireds are displayed in Figure 14. The amount of inter-syllable variation for the lower-central ired falls close to zero at the three points $t = 0.2, t = 0.5$, and $t = 0.7$. These

Figure 12: The mean accelerations of lip ireds along their respective principal axes of motion for "bob". The solid lines are for the lower-left, lower-central, and lower-right ireds, the dashed lines are for the upper lip ireds, and the dotted lines are for the right and left ireds.

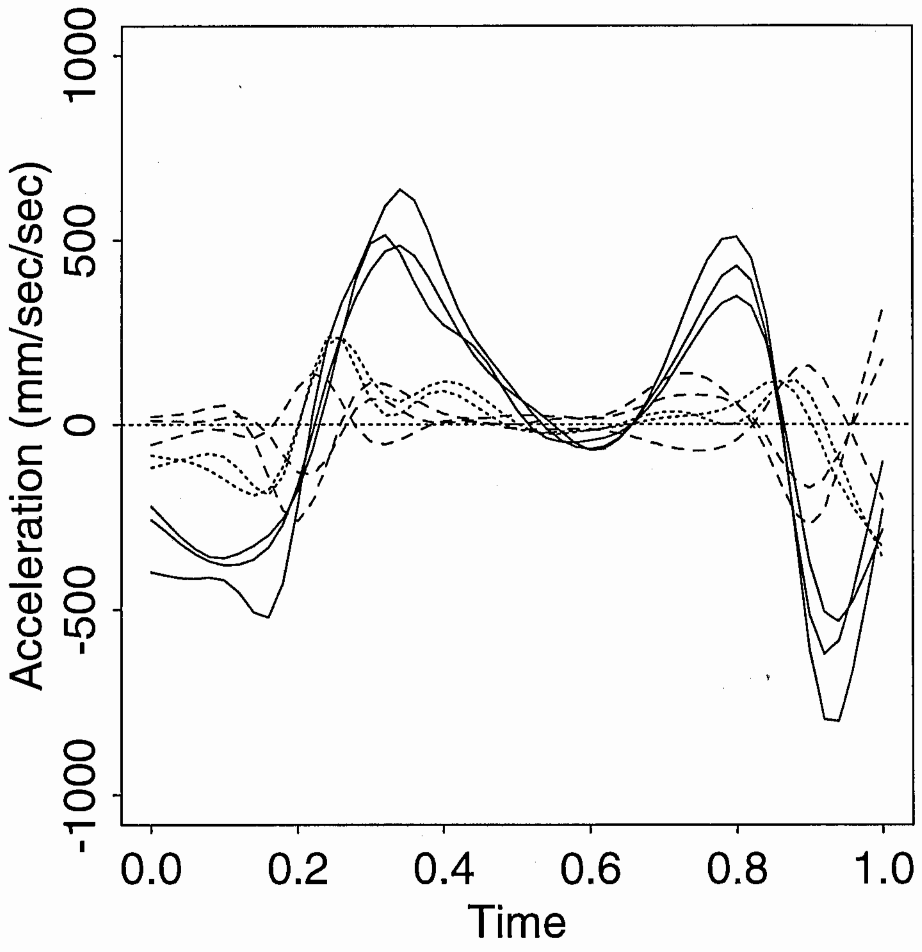
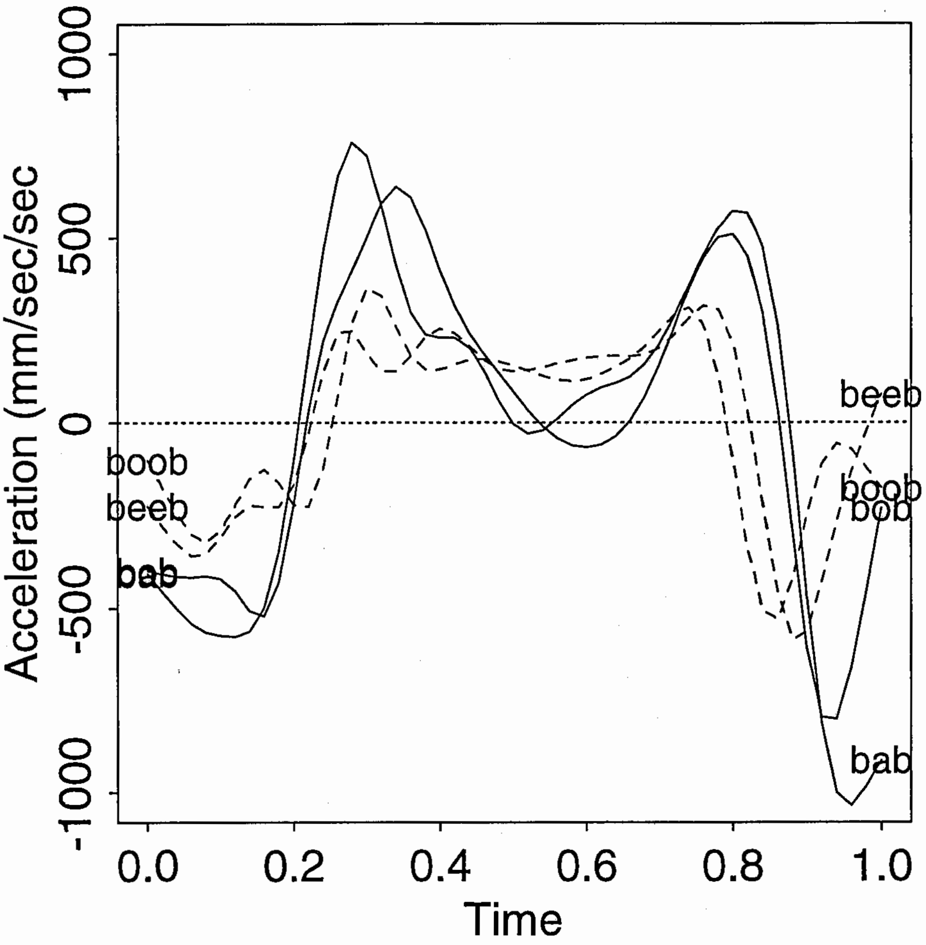


Figure 13: The mean accelerations of lip motion for all four syllables along their principal axes of motion for the lower-central ired. The solid lines are for the shorter syllables "bab" and "bob", and the dashed lines are for the longer syllables"beeb" and "boob".



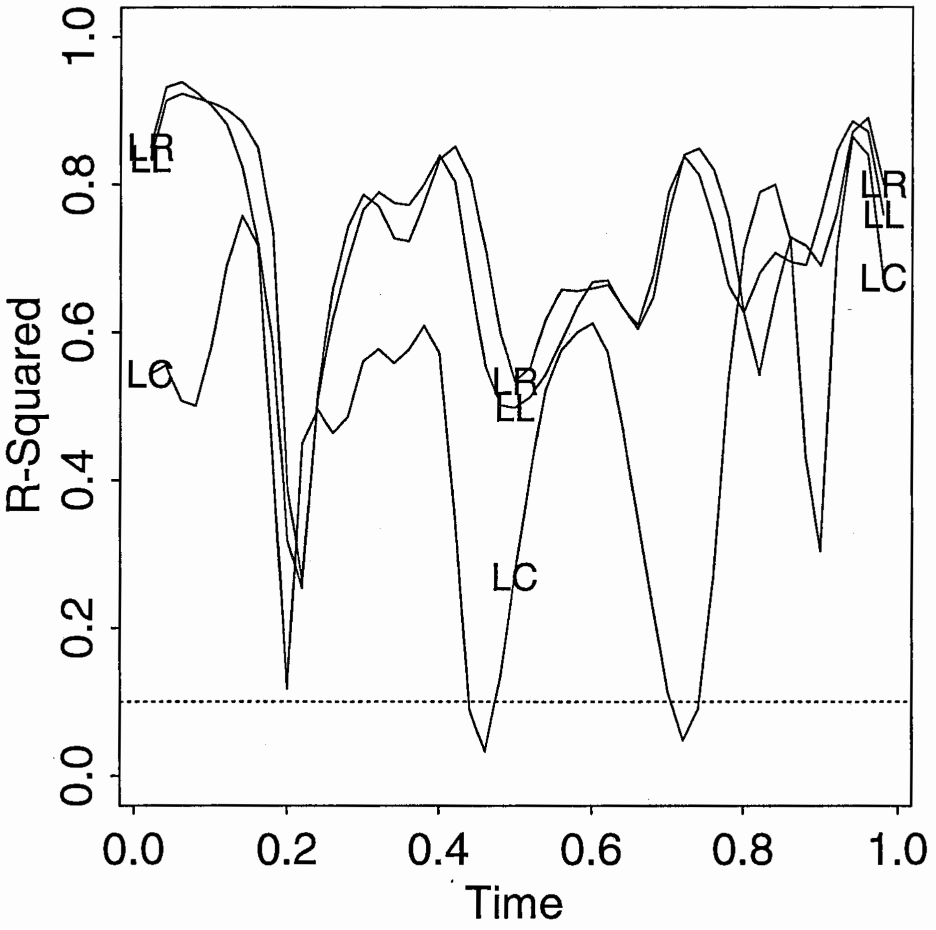
times precede the two points of strong acceleration and the stabilized point at $t = 0.6$. It would appear from this result that all four syllables may share a common timing process and a common strength of control.

The mean acceleration curves in Figures 13 and 14 indicate that the lower lip ireds have two brief episodes of strong positive acceleration bracketing a central period of zero acceleration, or in other words a central period in which lip movement is nearly linear. As with lip position, the mean acceleration curves for syllables show two groupings: one for "bab" and "bob" and the other for "beeb" and "boob". The latter two syllables show less of of a contrast between the two high-acceleration episodes and the central low acceleration phase.

Within-ired principal components analyses of accelerations revealed that variation in acceleration is much more complex than that for position since there were no clearly dominant components. The first four components for the lower-central ired, for example, accounted for 36.0, 25.2, 18.3, and 6.8 percent of the variation, respectively.

A cross-ired principal components analysis indicated that there three clearly dominant principal components, with the first four percents of variance accounted for being 28.8, 18.9, 14.0, and 8.4. In general, we found that

Figure 14: The squared correlation functions $R^2(t)$ for the functional analysis of variance of the lower lip ired variability across syllables.



the variation of the the acceleration patterns was more complex than the trajectories of lip positions, perhaps partly as a consequence of the lower statistical stability of the estimates of these curves.

6 Discussion

The study of speech articulation has been hindered by the sheer volume and complex time-varying character of data is available for analysis. Changes in the shape of the vocal tract are multidimensional in space and time, and exhibit functional variation within and across context. We have attempted to demonstrate how statistical techniques can be used to address questions such as the following: 1) How many degrees of freedom are exhibited by a single sensor? 2) How many degrees of freedom are exhibited by a set of sensors attached to a single articulator? These are theoretically important questions that can be addressed by functional principal components analysis. Although work has been carried out for static lip and tongue shapes (e.g., Linker, 1982), the present analysis addresses the variance in the movements themselves. 3) How many degrees of freedom are exhibited by a set of sensors attached to multiple articulators? This question asks whether articulators are acting as a motor synergy. Much recent work (e.g., Gracco, 1994; Munhall, Lofqvist and Kelso, 1994) has suggested that groups of articulators in speech act as coordinative structures with less degrees of freedom than the group could in principal show independently. To date the issue of functioanl organization has not been rigorously examined in speech production. 4. How does the behavior of a sensor, a group of sensors within an articulator, and a group of sensors across articulators change with context? This can be addressed with FANOVA as well as with a comparison of the different solutions to the principal components analyses. It is clear that different vowels have different lip shapes. However, it is a different question to ask whether the lips are coordinated differently to produce the movements for the two vowels. 5. How does the variance due to the endpoints of a segment differ from the variance intrinsic to the segments themselves? This is a fundamental question in the study of coarticulation and serial ordering. The use of spline interpolation offers new insights into this problem.

From the data analysis a number of tenative conclusions can be drawn. Lip motion for the speaker from whom the data were collected had a number of common characteristics for the four syllables used, and these are displayed in Figure 5. Most of the motion was in the three ireds placed on the lower lip.

As such,we were able to concentrate our attention for the most part on the lower-central ired. Moreover, this trajectory was sufficiently linear that we felt that little of importance was ignored by displaying and analyzing motion along the straight line that best fit this trajectory. This line was estimated by a principal components analysis of the mean motion within each ired, and we were able to use ired-specific local coordinate systems with the origin placed at the point defined by taking the position mean both spatially and temporally. It is clear from Figure 5, however, that these best local coordinate systems do vary substantially from syllable to syllable. For example, "boob' exhibited rather less retrusive movement than the other three syllables, and "bab"and "bob" showed greater vertical lower lip movement.

The spline smoothing technique used in these analyses served as much to partition the variation in lip position into end-effect motion and shape variation as it did to actually smooth the data. We felt that this enabled us to remove variation due primarily to where the lips were at the beginning and end of the syllable from variation in how the syllable was articulated. There are strong and significant differences between syllables in the movement of the ireds, both for the total motion and for shape alone,as displayed in Figures 5 and 9, and confirmed in the plots in Figure 10 of the R^2 fit measure for the functional analysis of variance. But these differences are still relatively small in magnitude relative to the amount of movement within any syllable.

Any conclusions drawn about the mechanisms controlling a system based solely on observations of the behavior of the system are necessarily tentative. Nevertheless, we found some indications that the lip motion associated with these syllables involved external or exogenous control in the central part of the syllable. The average lower lip motion seen in Figure 9 is clearly triphasic in character, with the opening and closing rapid movements being separated from the central seen in Figure 13 and the controlled linear movement in the center suggest an onset-control-offset cycle that is contained in terms of our arbitrary time units within the interval $0.3 \leq t \leq 0.8$,a period corresponding to about one-sixth of a second. We also noted that over the same period the correlation surface in Figure 8 was much flatter than in initial period, $t \leq 0.3$, and final period, $t \geq 0.8$. In order to get a clearer picture of the possible control processes involved, some covariate observations are desirable. We are currently investigating the use of jaw motion to separate motion purely internal to the lip tissue from that contributed by the jaw. We are also recording EMG activity in the relevant muscle groups.

In conclusion, we have attempted to indicate how statistical concepts can be extended to functional data based on the position that there is no fundamental distinction between data domains which are finite sets and those that are continua (Ramsay, 1982). Moreover, we have shown that treating movement data as functions has tremendous potential in understanding the organization and control of multidimensional movements. The techniques are robust and quantitative such that they can easily be adapted to many different measurements domains in which data vary as a function of time. Finally, the application of these techniques provided a principled approach to the reduction of the dimensionality of data rather that an arbitrary decision regarding the number of relevant variables to include or discard in any functional analysis.

REFERENCES

Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research, 37,* 4-27.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977) Factor analysis of tongue shapes. J. Acoust. Soc. Am. 6f2, 693-707.

Hirayama, M., Vatikiotis-Bateson, E., & Kawato, M. (1993) Physiologically-based speech synthesis using neural networks. *IEICE Trans. Fundamentals, E76-A,* 1898-1910.

Horn, B. K. P. (1987) Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America, 4,* 629-642.

Linker, W. (1982) Articulatory and Acoustic correlates of labial activity in vowels: A cross-linguistic study. *UCLA Working Papers in Phonetics, 56,* 1-134.

Munhall, K. G., Lofqvist, A, & Kelso, J. A. S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America, 95(6)*, 3605-3616.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika,* **47,** 379-396.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (with discussion.) *Journal of the Royal Statistical Society, Series B,* **53,** 539-572.

Schroeter, J. & Sondhi, M. (1992) Speech coding based on physiological models of speech production. In S. Furui & M. Sondhi (eds.) *Advances inSpeech Signal Processing.* Marcel Dekker: New York.

Wahba G. (1990) *Spline models for observational data.* Philadelphia: Society of Industrial and Applied Mathematics.

Table 1: The proportions of variance accounted for by the first three components of variation of within-in ired position along the first principal axis of motion for "bob".

Ired	I		II		III	
	Total	Shape	Total	Shape	Total	Shape
LR	71.6	88.5	13.5	7.5	9.4	2.9
LC	65.2	86.8	16.5	8.4	10.2	3.6
LL	67.4	86.4	16.1	8.6	9.5	3.8
L	85.0	77.6	6.5	15.4	5.9	4.2
UL	86.3	79.6	9.1	12.9	2.2	4.6
UC	87.1	70.5	8.4	12.3	1.9	11.3
UR	79.6	76.3	11.9	9.6	5.2	8.6
R	84.6	90.2	8.0	6.2	5.6	2.3