Functional data analyses of lip motion

J. O. Ramsay[a)]
McGill University, Montreal, Quebec H3A 1B1, Canada

K. G. Munhall
Queens University, Kingston, Ontario, Canada

V. L. Gracco
Haskins Laboratories, New Haven, Connecticut 06511

D. J. Ostry
McGill University, Montreal, Quebec H3A 1B1, Canada

(Received 18 December 1994; accepted for publication 1 February 1996)

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 techniques are described that provide useful tools for describing multivariate functional data such as the measurement of speech movements. The 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 variation within an articulator site from variation from between sites. 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, it is found that the analyses of acceleration were especially helpful in suggesting possible control mechanisms. The focus is on using these speech production data to understand the basic principles of coordination. However, it is believed that the tools will have a more general use.

PACS numbers: 43.70.Aj

INTRODUCTION

Two problems have complicated the study of the high-dimensional dynamic process of speech articulation. First, the movements of the vocal tract are spatially complex and there is significant motion in three dimensions, leaving the researcher with data that have many dimensions and therefore many degrees of freedom. While there have been some assessments of the dimensionality of static shapes of the lip (Linker, 1982) and tongue (Harshman *et al.*, 1977), there has been little work on the dimensionality of the motions of these articulators except for Maeda (1990).

Second, there is the problem of studying the dynamic or functional character of the process. One of the most common simplifying assumptions is to restrict the analyses to scalar summaries of the movement trajectories, such as the average duration, amplitude and peak velocity of markers attached to individual tissue points. These marker measures are then subjected to conventional univariate or multivariate statistical analyses. But using point summary measures of continuous functions presumes that these are sufficient for understanding the underlying process and, moreover, that the time-varying detail of the movement trajectories is relatively unimportant.

An alternative approach, called functional data analysis (FDA), has been developed by Ramsay and colleagues (Ramsay, 1982; Besse and Ramsay, 1986; Ramsay and Dalzell, 1991) in which the traditional multivariate analyses such as principal components analysis are expressed in functional analytic terms. Ramsay has demonstrated the utility of this approach by analyzing tongue movements in speech (Ramsay, 1982; Besse and Ramsay, 1986) and three-dimensional limb movements (Ramsay, 1989). FDA involves the definition of useful statistical analyses such as principal components analysis in functional analytic terms, and the variance components that are identified are functions. The modes of variation of trajectories are thus expressed in a form similar to the trajectories themselves.

This approach has clear advantages: (i) It takes account of the underlying continuity of the physiological system generating the behavior; (ii) it displays temporal dependencies in the data owing to this continuity; (iii) it provides methodologies to deal quantitatively with the complexities of multidimensional time series data like those collected in speech experiments; and (iv) functional data analysis offers the

[a)]Requests for preprints should be directed to the first author at Department of Psychology, 1205 Dr. Penfield Ave., Montréal, Québec H3A 1B1, Canada. Electronic mail: ramsay@psych.mcgill.ca

possibility of studying the variation among various orders of derivatives or linear combinations of derivatives. It may be, for example, that the significant modes of covariation across articulators may be at the level of acceleration, or, if the system is primarily harmonic, at the level of a specified linear combination of position and acceleration. Thus this approach can reveal temporal and spatial dependencies between articulators that are due to their shared patterns of motion.

In this study the OPTOTRAK system (for other uses of this technology see Bateson and Ostry, 1995) was used to measure lip motion. Three spatial coordinates of eight separate articular positions observed under four experimental conditions will be examined. The FDA techniques used in this paper deal with a partition or decomposition of data into fundamental components of variation. Principal components analysis serves to assess the complexity and dimensionality of across-replication variation in three-dimensional lip movement, taken both within and across recording positions. A functional version of analysis of variance (FANOVA) permits the study of across-condition variation in articulation.

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.

This paper is primarily methodological in that it aims to show functional data analysis in action within a context of high-dimensional dynamic data. Our aim is introduce these new statistical tools rather than to claim novel substantive results.

I. EXPERIMENTAL METHODS

A single subject, a male native speaker of Canadian English with no reported speech or language disorders, 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/. The subject spoke 20 repetitions of each syllable type, randomized across the four vowels.

The motion of the lips was monitored using OPTOTRAK, an optoelectronic tracking system that can transduce the 3-D position of markers. Eight infrared emitting diodes (ireds) were attached to the vermilion border of the lips using double-sided tape. An additional six ireds were positioned on a custom head-mounted jig in order 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. Three reference trials were collected prior to the experiment during which the subject held a Plexiglas jig between his teeth. Three ireds attached to the Plexiglas 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.

The observations for each syllable therefore consisted of 24 movement streams (8 ireds times 3 spatial dimensions) for each of 20 trials. 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 ms). To simplify data analysis, the data were interpolated so that each record had 51 equally spaced observations, and the time values 0,0.02,0.04,...,1 were assigned, and all results in this paper are given with respect to this artificial time frame.

II. 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, analyzes within and across marker locations in terms of their dominant or principal features. The first two techniques are discussed in more detail because accounts of them are not readily accessible in the applied statistical literature, but functional principal components, described elsewhere (Ramsay, 1982; Besse and Ramsay, 1986) is only summarized.

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^mx indicates the derivative or order m. A specific value of, say, acceleration at time t is indicated by $D^2x(t)$.

A. 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 spline smoothing procedure was developed especially for this application, and is therefore described at some length.

The basic idea behind spline smoothing (Eubank, 1988; Green and Silverman, 1994; Wahba, 1990) 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. The spline smoothing criterion for assessing the fit of smoothing function x_i for replication i of coordinate X used in this paper is

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

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 term 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 emphasis on the penalty $\int[D^4x(t)]^2\ dt$ in the second term, and therefore the more fit that must be sacrificed in order to keep this term comparable in size to the first. It is instructive to consider the two limiting cases. As $\lambda\to\infty$ the size of the fourth derivative is ultimately forced to zero. This implies that the fitted function x would become a cubic polynomial, for which $D^4x=0$ exactly. At the other extreme, as $\lambda\to 0$, 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. The idea behind this strategy is to consider what would happen if the fitted function x were fit to all but the kth curve value, and then this actual curve value were compared to the predicted value. Conceptually this approach could be applied 51 times per curve by leaving each observation out in turn. Finally the squared errors of prediction could be accumulated to provide a global measure of lack of fit. This technique is called cross validation. The smoothing parameter λ would then logically be chosen to so as to minimize this cross-validated error sum of squares. However, in practice, the cross-validation approach can be prohibitively time-consuming, and the generalized cross-validation method involves some short cuts to approximate the consequences of a complete cross validation, while retaining the speed of a single smoothing step.

Smoothness is assessed 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.

B. 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:

(1) end-point variation, or variation near the ends of curves, and
(2) 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 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 is split up as follows:

$$x(t)=u(t)+e(t), \qquad (2)$$

where

(1) 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 *end-point* component of x.

(2) e is the function that has values and derivative values equal to 0 at the end points, 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.

C. Functional analysis of variance

We shall need to explore the systematic differences among lip position functions x, as well as their acceleration counterparts, across the four experimental syllables. 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.

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 $x_{ij}(t)$, $y_{ij}(t)$, and $z_{ij}(t)$ and their derivatives, subscript j indexing syllable. This implies the counterpart functional ANOVA, or FANOVA model

$$x_{ij}(t)=\mu(t)+\alpha_j(t)+e_{ij}(t) \tag{3}$$

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

$$\sum_j \alpha_j(t)=0 \quad \text{for all} \quad t. \tag{4}$$

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_j one proceeds as follows. Making use of the fact that the sample size $N=20$ is the same for each syllable, and indicating the number of conditions by $J=4$, the parameter estimates are

$$\begin{aligned}\hat{\mu}(t)&=(JN)^{-1}\sum_i \sum_j x_{ij}(t),\\ \hat{\alpha}_j(t)&=(N)^{-1}\sum_i x_{ij}(t)-\hat{\mu}(t).\end{aligned} \tag{5}$$

Residual functions $\hat{\epsilon}_{ij}$ are then estimated by

$$\hat{e}_{ij}(t)=x_{ij}(t)-\hat{\mu}(t)-\hat{\alpha}_j(t).$$

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

$$SSE(t)=\sum_i \sum_j [\hat{e}_{ij}(t)]^2.$$

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

$$F(t)=\frac{[SSE_0(t)-SSE(t)]/(J-1)}{SSE(t)/[N(J-1)]},$$

where SSE_0 is the null hypothesis error sum of squares,

$$SSE_0(t)=\sum_i \sum_j [x_{ij}(t)-\hat{\mu}(t)]^2.$$

For a fixed value of $t, F(t)$ has, in this application, numerator and denominator degrees of freedom 3 and 76, respectively.

D. 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.

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. This technique has been in use for some time, and Ramsay (1982) and Besse and Ramsay (1986) can be consulted for further technical details.

III. EXAMPLE ANALYSES

In this section we will summarize a series of analyses of the ired motions. Our aim is to demonstrate that FDA allows the researcher to explore questions about speech motor control that are not easily accessible through more traditional analyses of ired positions at selected points in time. Some parts of an FDA approach coincide with the standard repertoire of speech analyses. We begin with some descriptive analyses that will share many features with standard point analyses.

A. Descriptive displays and analyses

Figure 1 shows the three coordinate functions for three of the ireds positioned at the upper center, lower center and extreme right, 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-Z or sagittal plane 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.

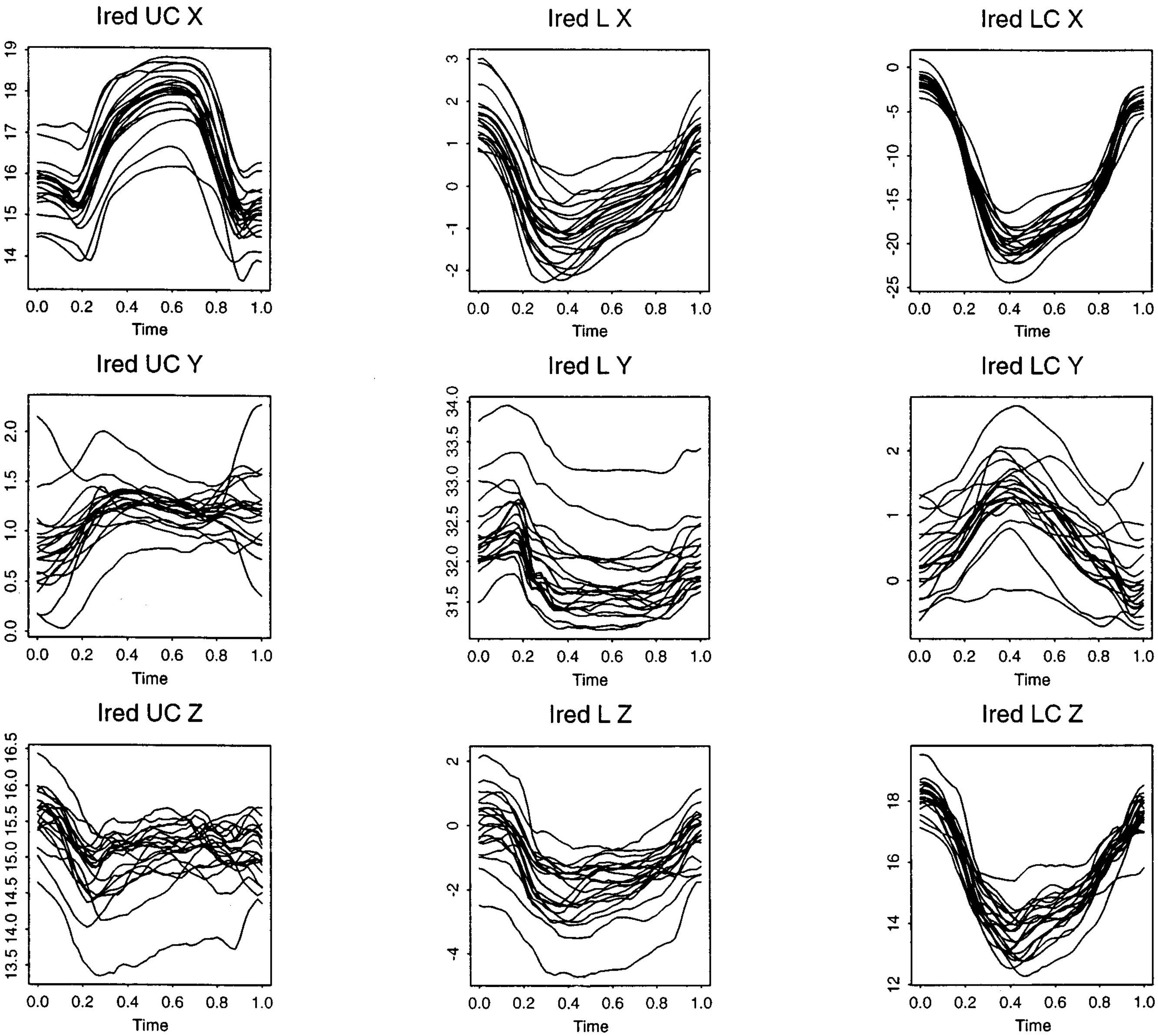

FIG. 1. Twenty records for X, Y, and Z coordinates for infrared emitting diode positions at the upper-center, extreme left and lower-center, respectively, during the utterance of the syllable /bab/. The X direction is vertical position, the Y direction is lateral position, and the Z direction is protrusion or fore/aft position, all in millimeters. Note that the scale of the ordinate is different for each panel.

Figure 2 displays the mean movement of all eight ireds for each syllable in the sagittal plane. The syllables /bæb/ and /bab/ involve greater movement of the lower lip than /bib/ and /bub/, and /bub/ involves more forward or protruding lower lip movement than the others. A number of observations could be made from these average trajectories. First the vowels differ in the magnitude of the movements involved in their production. The vowels in /bæb/ and /bab/ have larger movements than /bib/ and /bub/. Second the /bub/ trajectories differ from the other three vowels because of the rounding for that vowel. Finally the motions are quite simple and are generally linear.

B. Principal axis transformation

The fact that within-ired motion is nearly linear suggests that the 3-D 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 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 three variance-covariance matrix for ired coordinates, and transforming the 51 times three 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 mo-

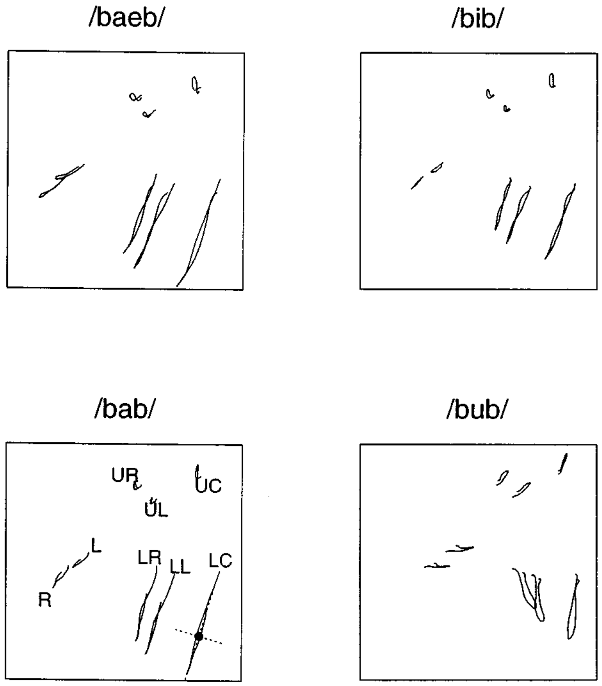

FIG. 2. Sagittal plane mean motion for each syllable and all for ireds. The ireds are identified in the plot for /bab/, and this plot also indicates the two principal axes of variation for the lower lip ired by dashed line segments. The axis scales for all plots are the same, and omitted for clarity.

tion, 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 Fig. 2. The motion of this ired along the first principal axis is displayed in Fig. 3. One notes that motion of the lower central ired along the first principal axis

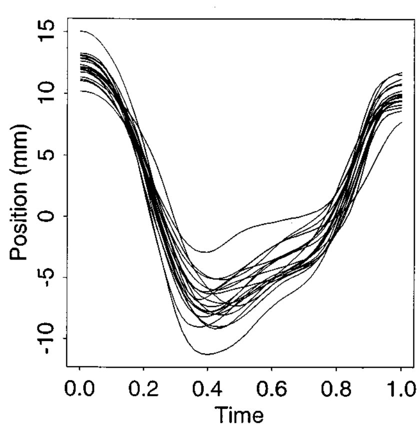

FIG. 3. The motion of the lower central ired along the principal and dominant axis of motion for /bab/ relative to the mean position.

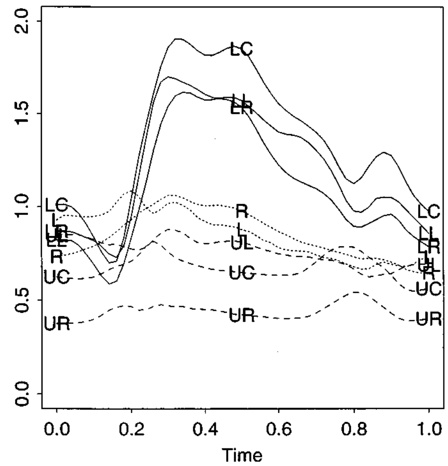

FIG. 4. Standard deviation of lip positions for all eight ireds during /bab/ along each ired's first and dominant principal axis. Curves for the lower right, central and left ireds are solid lines, for the upper right, central and left are dashed, and for the extreme left and right are dotted.

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.

Thus PCA of the lower-central trajectories confirms that variation in individual curves corresponds to what is evident by inspection in the mean trajectories in Fig. 2, namely that lower-central lip motion is primarily one dimensional in character.

C. Correlation analyses

Trajectories can reveal a great deal about the underlying control mechanism (Atkeson and Hollerbach, 1986). One way to examine the different influences on articulator motion is to evaluate, in a functional sense, the standard deviations and the correlations within and among the ireds. The standard deviation curves in Fig. 4 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 function of the two time values, these correlation values can be displayed as a surface over the time by time plane.

Figure 5 shows the correlation surfaces for movement along the first principal component of movement for the lower-central ired as a perspective plot of the surface. The

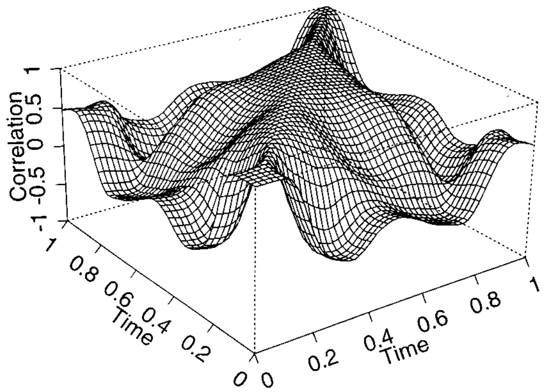

FIG. 5. The correlations among lip positions along the dominant principal component of movement for the lower center ired are displayed as a perspective surface plot. The diagonal ridge running from front to back contains the unit or near unit correlations for pairs of very similar time values. Note the flat region just after the middle of the interval.

diagonal ridge running from foreground to background contains the unit correlations for equal time values.

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 toward zero rather rapidly, 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 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 Fig. 5 with Fig. 4 indicates that covariance in this region is elevated relative to the standard deviation. These flat regions in the correlation surface, then, could indicate that the system is under external or exogenous control. Similar effects were noted in Ramsay (1982).

D. Functional analysis of variance

It is clear from the sagittal plane plot in Fig. 2 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, that motion along these axes tends to differ systematically from syllable to syllable. The left part of Fig. 6 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 /bæb/ than for /bub/ for this ired. One notes two nodes where all four trajectories tend to coincide. The right part of Fig. 6 displays only the shape effect, and gives a better idea of how the trajectories differ once end position differences are removed. Syllables /bib/ and /bub/ not only show less motion than the other two, but also exhibit less asymmetry in their trajectories.

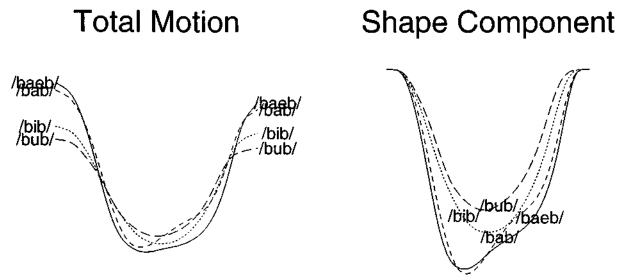

FIG. 6. The average position of the lower central ired along the principal axis of motion is plotted for each syllable in the left display. In the right display only the shape component is plotted.

In order to confirm that the differences such as those noted in Fig. 6 are substantial in a statistical sense, functional analysis of variance of the motion of each ired along that ired's principal axis of motion can be carried out, for the total motion and for the shape component. The strength of the intersyllable variation is summarized in Fig. 7 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 intersyllable 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 intersyllable variation in the center of the interval as well. The shape components, however, have substantial variation over the rest of the interval, including at the two acceleration episodes.

E. 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

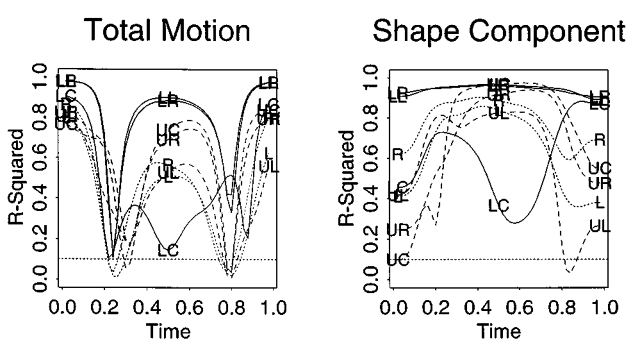

FIG. 7. The left display of shows the squared multiple correlation R^2 as a function of time for the analysis of between-syllable variance for all ireds and for total motion. The right display is for shape component only. Curves for the lower right, central and left ireds are solid lines, for the upper 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 76 degrees of freedom is shown as a dotted horizontal line.

TABLE I. 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 /bab/.

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

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 I gives the proportions of variances accounted for by the first three principal components for /bab/. 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 I, where motion along the single dominant trajectory accounted for about 87% of the variance. The primary type of variation was simply how wide the lips were opened, with the shape of the trajectory remaining relatively unchanged.

The first principal 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, but a collection of variables can exhibit quite complicated and high-dimensional patterns of combined variation.

The proportions of variance accounted for by the across-ired principal components are displayed in Fig. 8 for both total curve variation and for shape variation alone. Note that although there are in principal as many as 24 substantial principal components possible, the sample size of 20 imposes the actual limit of 19. However, the principal components analysis of simultaneous ired motion, especially when only shape components were used, suggested strongly that

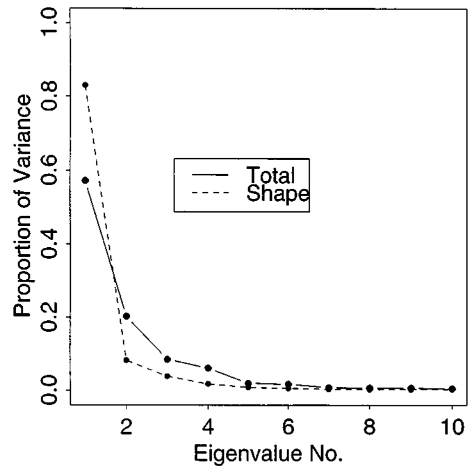

FIG. 8. The proportions of variance accounted for by the first ten principal components of simultaneous or joint ired variation along each ired's first principal axis for /bab/.

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 /bab/.

F. Displays and analyses of lip acceleration

It would seem also profitable to look closely at the acceleration of the system since forces resulting from muscle contraction and inertial loads applied to the lip tissue will, by Newton's third law, generate immediate changes in acceleration, while the same forces will affect position, being two integrals removed from acceleration, only gradually. For these and other reasons, it seems plausible that the various controlling processes determining lip motion will have their most visible impact on acceleration. The spline smoothing technique adopted in Sec. I B was chosen in order to give smooth and accurate estimates of acceleration, although the technique does not separate shape from end-effect variation in acceleration, as it does for position.

Figure 9 shows the average acceleration functions for each of the ireds along their respective principal axes of variation for /bab/. The lower coordinates (solid lines) are first accelerated negatively or downward ($t \leq 0.2$), and then pass through a positive acceleration phase ($t=0.3$) during which the descent of the lower lip is stopped. This lip opening phase is followed by a short period of near zero acceleration ($t=0.5$) corresponding the period of slow and linear change in Fig. 3, followed by another strong acceleration upward initiating lip closure ($t=0.8$), and then finally completed by a negative acceleration episode as the lip returns to the closed position. The other ireds show less acceleration and more complex patterns.

Figure 10 indicates the variation in the lower central ired acceleration across syllables in their respective first principal

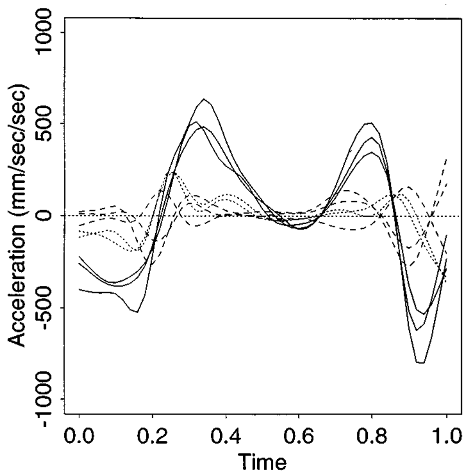

FIG. 9. The mean accelerations of lip ireds along their respective principal axes of motion for /bab/. The solid lines are for the lower-left, lower-central, and lower-right ireds, the dashed lines are for the upper lip ireds, and the dotted lines are for the right and left ireds.

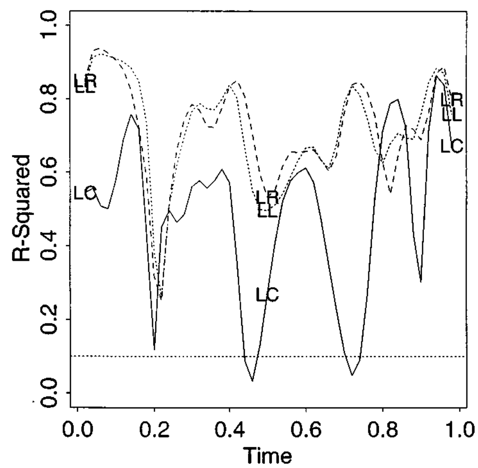

FIG. 11. The squared correlation functions $R^2(t)$ for the functional analysis of variance of the lower lip ired acceleration variability across syllables.

axes of motion. The longer syllables, /bib/ and /bub/, fail to exhibit the momentary period of zero acceleration of the shorter syllables, and exhibit less acceleration or smoother motion throughout their trajectories. This corresponds to the more rounded contours of their position functions in Fig. 6.

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 Fig. 11. The amount of intersyllable 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 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.

Within-ired principal components analyses of accelerations revealed that variation in acceleration is rather 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.

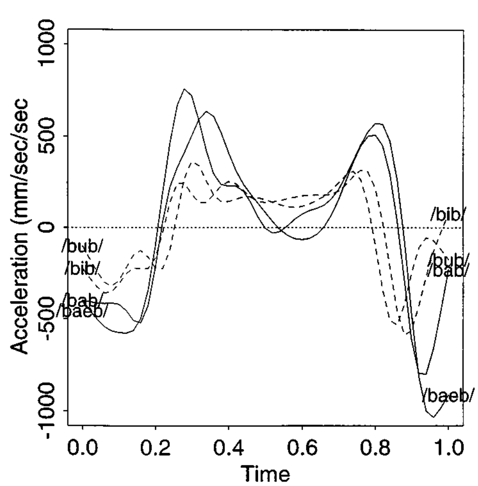

FIG. 10. The mean accelerations of lip motion for all four syllables along their principal axes of motion for the lower-central ired. The solid lines are for the shorter syllables /bæb/, and /bab/, and the dashed lines are for the longer syllables /bib/ and /bub/.

A cross-ired principal components analysis indicated that there were three clearly dominant principal components, with the first four percents of variance accounted for being 28.8, 18.9, 14.0, and 8.4 for a total of 70.1%. In general, we found that the variation of 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.

IV. DISCUSSION

The study of speech articulation has been hindered by the sheer volume and complex time-varying character of data. The first task of the functional data analytic tools applied in this paper was to reveal the simplicity underlying the potentially complex motions of eight lip positions in 3-D space. Principal components analysis of within-ired motions displayed the essentially straight-line trajectories of all ireds (Fig. 2), permitting us to replace the three X, Y, and Z coordinates by a single coordinate indicating position along these trajectories. While it was convenient that the trajectories turned out to be linear, single coordinates for curvilinear trajectories identified in this way can also be constructed with some additional effort.

Much recent work (e.g., Gracco, 1994; Munhall *et al.*, 1994) has suggested that groups of articulators in speech act as coordinative structures with less degrees of freedom than the group could potentially show independently. An across-ired principal components analysis provided further confirmation by showing that ireds operate in concert to a rather impressive degree, so that their covariation can be well-summarized by the same number of principal components as required to summarize the variation of any single ired (Fig. 8). In this way the multiple potential dimensions of variation were reduced to the few important dimensions of variation of the lower-central ired along its principal component trajectory.

The theme of data reduction was also worked out in the context of single curves. The spline approximation technique was customized for this application to separate end-effect variation from interior or shape variation, and this also reduced the complexity of variation across time by showing that there was one less principal component for shape than for the total curves (Table I).

We were also interested in syllable effects, so that after these preliminary data-reduction analyses it was possible to show via functional analysis of variance that there are important sources of intersyllable variation, as evidenced in Figs. 7 and 11. Moreover, the functional nature of the resulting $R^2(t)$ effect sizes highlighted the fact that intersyllabic effects are strongly time variant, being nearly negligible at the two episodes of strong positive acceleration and large between.

Finally, although these data can offer only a few hints about the underlying control mechanisms generating these effects, we found that the correlation surface plot Fig. 5 was suggestive of a triphasic process, with the nature of the control being markedly different in the central third of the process than near the ends. This was also clear in the various analyses of acceleration, and we conjecture that future investigations will find much of interest in an exploration of second derivative information. Getting a good estimate of a second derivative in the presence of even a small amount of noise is a statistically challenging task, and the spline smoothing procedure described here is especially appropriate in that it yields a smooth acceleration function.

The main goal of this paper has been to showcase the power of functional data analytic techniques, and to describe in some detail how to use them. It is acknowledged, on the other hand, that some of the results obtained here could have been obtained by applying more conventional analysis. This is certainly true of the principal components analyses, and the novelty of the functional ANOVA results lies primarily in the graphical displays that result. On the other hand, the spline smoothing and decomposition process has no counterpart in more standard statistical analyses, and played a central role. In particular, the possibility of analyzing acceleration and other types of derivative information is, in our view, the most promising aspect of the functional approach.

From a substantive point of view, it would be clearly desirable to be able to remove jaw motion and its effects from lip motion, and our current investigations are doing this. If we are to learn more about control processes and their relation to lip position, velocity and acceleration, muscle activity measures such as EMG records are essential. And, of course, more than 20 replications would imply more stable estimates of effects as well as the possibility of identifying other more subtle aspects of variation.

We have learned from these analyses, in addition, that future experiments can be simplified in important ways. Eight lip positions can probably be reduced to two or three at most, and motion in the sagittal plane, at least for central articular positions, is sufficient for most purposes. Very high sampling rates are not required for reasonably accurate estimates of first and second derivatives when the noise level is of the order of that in OPTOTRAK measures.

ACKNOWLEDGMENT

This research was supported by NIH Grant No. DC-00594 from the National Institute on Deafness and other Communication Disorders.

Atkeson, C. G., and Hollerbach, J. M. (**1985**): ''Kinematic features of unrestrained vertical arm movements,'' J. Neurosci. **5**, 2318–2330.

Bateson, E. V., and Ostry, D. J. (**1995**). ''An analysis of the dimensionality of jaw movement in speech,'' J. Phon. **23**, 101–117.

Besse, P., and Ramsay, J. O. (**1986**). ''Principal components analysis of sampled functions,'' Psychometrika **51**, 285–311.

Eubank, R. (**1988**). *Spline Smoothing and Nonparametric Regression* (Dekker, New York).

Gracco, V. L. (**1994**). ''Some organizational characteristics of speech movement control,'' J. Speech Hear. Res. **37**, 4–27.

Green, P. J., and Silverman, B. W. (**1994**). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (Chapman and Hall, London).

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). ''Factor analysis of tongue shapes,'' J. Acoust. Soc. Am. **62**, 693–707.

Horn, B. K. P. (**1987**). ''Closed-form solution of absolute orientation using unit quaternions,'' J. Opt. Soc. Am. **4**, 629–642.

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

Maeda, S. (**1990**). ''Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,'' in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, New York).

Munhall, K. G., Lofqvist, A., and Kelso, J. A. S. (**1994**). ''Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip,'' J. Acoust. Soc. Am. **95**, 3605–3616.

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

Ramsay, J. O. (**1989**). ''The analysis of replicated spatial functions of time,'' in *Multiway Data Analysis*, edited by R. Coppi and S. Bolaxco (North-Holland, Amsterdam), pp. 363, 374.

Ramsay, J. O., and Dalzell, C. J. (**1991**). ''Some tools for functional data analysis (with discussion),'' J. R. Stat. Soc. Ser. B **53**, 539–572.

Wahba, G. (**1990**). *Spline Models for Observational Data* (Society of Industrial and Applied Mathematics, Philadelphia).