The Historical Functional Linear Model

N. MALFAIT & J. O. RAMSAY
Department of Psychology, McGill University,
Montreal H3A 1B1, Quebec, Canada
malfait@steve.psych.mcgill.ca
ramsay@psych.mcgill.ca

S. FRODA
Department of Mathematics, Université du Québec à Montréal,
Montreal H3C 3P8, Quebec, Canada
froda@math.uqam.ca

October 12, 2001

Summary

We develop a functional linear model where the values at t of a sample of curves $y_i(t)$ are explained in a feed-forward sense by the values of covariate curves $x_i(s)$ observed at times $s \leq t$. Special attention is given to the case $s \in [t-\delta, t]$, where the lag parameter δ is estimated from the data. The finite element method is used to estimate the bivariate parameter regression function $\beta(s,t)$, which is defined on the triangular domain $s \leq t$. The model is applied to the problem of predicting the acceleration of the lower lip during speech on the basis of electromyographical recordings from a muscle depressing the lip. Simulation results are also provided to guide the calibration of the fitting process.

Key words: Finite element method; functional data analysis; functional linear model.

1 Introduction

This paper considers a functional regression problem in which one curve is used as the independent variable to explain the variation in another. Ramsay and Silverman (1997) considered several such models, but here we look at a new situation in which the influence of the carrier function on the outcome process is of a feed-forward nature. The model proposed here is useful in statistical analyses involving longitudinal, functional or time series samples, which are widely used in natural, medical and social sciences.

As an illustration, consider that an indicator of a patient's recovery, $y(t)$, may depend linearly on the time course of a treatment variable $x(s)$, and that this relation logically only involves times $s \leq t$. Moreover, there may be some reason to suppose that only treatments at times $s \in [t - \delta, t]$ for some lag $\delta \geq 0$ are likely to have an impact on $y(t)$.

In this paper, we apply this model to data from a speech production experiment. In Sections 2 and 3 we describe the data and what we call the historical linear model. In Section 4 we define least squares estimation for the regression function in the linear model, and in Section 5 we define the triangular finite element basis that we use for its expansion. Techniques for assessing fit are taken up in Section 6, and the results of applying the method to the speech data in Section 7.

2 The Speech Data

Speech is what our species does best. In conversation, an English speaker easily pronounces 14 phonemes per second, and this rate appears to be limited by the cognitive aspects of language rather than by the physical ability to perform the articulatory movements. Considering the muscles of the thoracic and abdominal walls, the neck and face, the larynx and pharynx, and the oral cavity, there are over 100 muscles that must be controlled centrally (Lenneberg, 1984).

The timing of the activation of different groups of muscles is a central issue for the anatomy and physiology of speech. A noninvasive method for collecting data on muscle activation comes from electromyography (EMG), which exploits the fact that muscle contractions are accompanied by electrochemical changes. A resting muscle is isopotential and generates no current, but when it is stimulated, the resulting excitation travels along the muscle as waves of action potentials; and if electrodes are attached to the skin over the muscle, they can pick up these potentials. The time taken for a neural signal to be transduced into muscle contraction can vary considerably, but for the facial muscles a representative value is about 50 msec.

Figure 1 displays the data that are analyzed in this paper. A subject was required to say the syllable "bob" $N = 32$ times. The duration of the syllable varied, but was time-normalized to 700 msec. The movement of the center

of the lower lip is shown in the upper panel. The middle panel shows the accelerations of the center of the lower lip, denoted as $y_i(t), i = 1, \ldots, N$, and accelerations, by Newtons's second law, reflects the force applied to tissue by muscle contraction. The lower panel displays the so-called linear envelopes of EMG activities, denoted by $x_i(s)$, recorded in the primary muscle depressing the lower lip, the depressor labii inferior (DLI). More information about the preprocessing steps will be found in the appendix.

The lower lip trajectory can be segmented into roughly five epochs, separated by dotted lines in Figure 1. The central interval corresponds to the /o/, during which the lip is stationary. To produce each /b/, the lip moves up and down. The DLI muscle plays two roles: Antagonist, when it brakes the movement during the ascending phases, and agonist when it accelerates the lip during the descending phases. Antagonist episodes are reflected by EMG bursts as the acceleration crosses zero while moving from a positive to a negative phase; and agonist activity by bursts at the start of the lip's descents.

3 The historical linear model

Let 0 and T indicate the initial and final times of the 32 records, and let δ indicate a time lag, beyond which we conjecture that there is no feed-forward type influence of $x(s)$ on $y(t)$. That is, $y(t)$ is influenced by $x(s)$ for

$s_0 \leq s \leq t$, with $s_0 = \max(0, t - \delta)$.

We will assume that $x(s)$ influences $y(t)$ linearly according to the following model that integrates this influence from s_0 to t:

$$y_i(t) = \alpha(t) + \int_{s_0}^{t} x_i(s)\beta(s,t)\,ds + \epsilon_i(t) \ , \qquad t \in [0,T]. \tag{1}$$

Here the function $\alpha(t)$ is a fixed intercept function that allows for the relationship between the mean lip and mean EMG curves, but cannot accommodate their covariation effects. The residual function $\epsilon_i(t)$, reflects the inability of the linear model to fit the data completely, and we assume that $\mathrm{E}[\epsilon_i(t)] = 0$ with $\mathrm{Cov}[\epsilon_i(t), \epsilon_j(t)] = 0, \ i \neq j$.

We might call this the *historical linear model* in the sense that the influence of $x(s)$ is assumed to be only forward in s, and therefore is part of the history of $y(t)$. Hastie and Tibshirani (1993), Ramsay and Silverman (1997) and others have considered the pointwise model, also called the varying coefficient model, resulting from $\delta \to 0$,

$$y_i(t) = \alpha(t) + x_i(t)\beta(t) + \epsilon_i(t) \ , \tag{2}$$

in which the regression function $\beta(t)$ depends only on t. The central question in this research is, then, whether β depends on both s and t, as opposed to only t, and how far back, δ, this dependency goes. Since $s \leq t$, the regression coefficient function $\beta(s,t)$ is defined on a subset of a triangle, as illustrated in Figure 2.

A model involving a bivariate regression function has already been considered by Ramsay and Silverman (1997). Model (1) is a refinement of this previous model which integrates over all of $[0, T]$. Here we modify the shape of the domain of β from square to triangular, to reflect the asymmetry of the relation between the two sets of functions. Also, the domain restriction in model (1) is intended to improve explanative and descriptive qualities, in contrast to predictive ones.

4 Estimation of the regression function $\beta(s,t)$

Let us first simplify model (1) by dropping the intercept function $\alpha(t)$. Since, from the normal equations,

$$\alpha(t) = \bar{y}(t) - \int_{s_0}^{t} \bar{x}(s)\beta(s,t)\, ds \ ,$$

by substitution, we obtain

$$y_i^*(t) \ = \ \int_{s_0}^{t} x_i^*(s)\beta(s,t)\, ds + \epsilon_i(t) \tag{3}$$

where $x_i^*(s) = x_i(s) - \bar{x}(s)$. In order to further simplify expressions, we drop the asterisk in what follows.

Regression function $\beta(s,t)$ can be approximated by an expansion $\widetilde{\beta}(s,t)$ in terms of K known basis functions

$$\widetilde{\beta}(s,t) = \sum_{k=1}^{K} b_k \phi_k(s,t) \ . \tag{4}$$

Defining

$$\psi_{ik}(t) = \int_{s_0}^{t} x_i(s)\phi_k(s,t)\,ds \ , \tag{5}$$

we have the alternative formulation

$$\begin{aligned} y_i(t) &= \sum_{k=1}^{K} b_k \int_{s_0}^{t} x_i(s)\phi_k(s,t)\,ds + \int_{s_0}^{t} x_i(s)\epsilon_a(s,t)\,ds + \epsilon_i(t) \\ &= \sum_{k=1}^{K} b_k \psi_{ik}(t) + \epsilon_i'(t) \ , \end{aligned} \tag{6}$$

where $\epsilon_a(s,t) = \beta(s,t) - \widetilde{\beta}(s,t)$ is the approximation error, and $\epsilon_i'(t)$ combines the random and the approximation errors.

5 The triangular finite element basis

We use the *finite element method*, often used to solve partial differential equation systems, to represent $\beta(s,t)$ as a continuous piecewise linear function constructed from piecewise linear basis functions that are zero everywhere except for a small region (Ames, 1992; Brenner & Scott, 1994). First the domain of $\beta(s,t)$ is subdivided into triangular regions, as shown in Figure 2. Sixteen triangles are shown in the figure, but in fact we used up to 676 triangles in our work. These triangles are the so-called the *elements* of the method.

The vertices of these triangles, called *nodes*, correspond to the basis functions $\phi_k(s,t)$ as follows. For each node k, $\phi_k(s,t)$ is linear in s and t over all triangles, continuous, equal to 1 at the node itself, and zero over all triangles

not having this node as a vertex. For nodes not on the boundary of the domain, these conditions imply that $\phi_k(s,t)$ is nonzero only over the hexagonal region made up of triangles next to node k. Figure 3 shows the tent-like form of the basis function for node 13. These conditions also imply that, at any point (s,t) in the interior of a given triangle, there will be at most only three basis functions having nonzero values, and these will correspond to the three nodes that are the vertices of this triangle.

This choice of a basis system has a number of advantages over the usual practice of representing bivariate functions using tensor product basis functions, such as tensor product B-splines. B-spline basis functions are defined over a rectangular grid, equivalent to using rectangular elements. Rectangular elements do not conform naturally to the triangular domain needed here, whereas triangular elements do. Moreover, the number of basis functions having positive values for a given t is determined by counting the number of nodes next to the corresponding line of integration. For example, with the triangulation shown in Figure 2, this number is seven for $t = 0.43$. Then, as the number of triangular elements increases, the proportion of basis functions having positive values for a given t decreases. As a consequence the design matrix in (7) below defining estimated values for the b_k's, which is sparse, becomes increasingly so as the discretization mesh is refined. The use of sparse matrix computational methods can greatly speed up the solution of

this equation. Finally, the evaluation of any basis function over an element involves the evaluation of a function linear in s and t, and therefore is very fast.

The triangulation shown in Figure 2 is also helpful when considering how large the lag δ should be in modelling the feed-forward influence of $x(s)$. Triangles falling beyond δ units from the diagonal are simply eliminated, so that the manipulation of δ corresponds to selecting subsets of the basis functions. Of course, we can only set δ at discrete values, but this is not a problem if we make the triangular mesh sufficiently fine. Letting λ indicate the width of a single triangle, we are permitted to use lag values $\delta = m\lambda$, for integers $m = 0, \ldots, M$ ($M\lambda = T$); and the number of nodes is then $(m+1)(M - m/2 + 1)$. When $\delta = m = 0$, $\widetilde{\beta}(s,t)$ reduces to a polygonal spline representation of the variation on the diagonal $\beta(t,t)$ with $M+1$ basis functions. Note also that the use of a sufficiently fine triangulation permits us to assume that the $\epsilon_i'(t)$ are globally homoskedastic, even if they are not expected to be so over one single element.

We may approximate the model in the form (6) by a multivariate linear model as follows. Evaluating $y_i(t), i = 1, \ldots, N$ at a finite set of time points $t_q, q = 0, \ldots, Q$, for each record i leads to

$$E[\mathbf{Y}_i] = \mathbf{\Psi}_i \mathbf{B} \ , \tag{7}$$

where

$$\mathbf{Y}_i = \begin{bmatrix} y_i(t_0) \\ \vdots \\ y_i(t_Q) \end{bmatrix}, \qquad \mathbf{\Psi}_i = \begin{bmatrix} \psi_{i1}(t_0) & \cdots & \psi_{ik}(t_0) & \cdots & \psi_{iK}(t_0) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \psi_{i1}(t_Q) & \cdots & \psi_{ik}(t_Q) & \cdots & \psi_{iK}(t_Q) \end{bmatrix},$$

and $\mathbf{B} = (b_1, \ldots, b_K)'$.

By stacking these matrices $\mathbf{Y}_i$ and $\mathbf{\Psi}_i$ on top of each other to obtain $N(Q+1) \times 1$ and $N(Q+1) \times K$ matrices $\mathbf{Y}$ and $\mathbf{\Psi}$, respectively we obtain the least squares estimator of $\mathbf{B}$,

$$\widehat{\mathbf{B}} = (\mathbf{\Psi}'\mathbf{\Psi})^{-1}\mathbf{\Psi}'\mathbf{Y} \ . \tag{8}$$

We found that $Q = 4n$ equally spaced time points, where n is the number of intervals in $[0, T]$ used for the triangulation, gave a satisfactory level of precision, with acceptable computational overhead.

6 Assessing Fit

The assessment of fit by comparing two historical linear models, one embedded within the other, raises a number of issues, and our proposals in this section are only intended to be preliminary to further work. However, the historical linear model does share with all linear models the property of mapping a parameter vector space, denoted by G_1, of dimension df_{G1} into an outcome Hilbert space H of dimension df_H by means of a linear operator L. The image of this mapping is denoted by $\hat{Y}_1$. The space G_1 here consists of regression functions $\beta(s, t)$ representable by the finite element basis, the

dimensionality df_{G1} of which equals the number of nodes determined by the mesh density M and lag $\delta = m\lambda$. A model assessment involves comparing a fit based on all of G_1 with that provided by a subspace G_0 of dimensionality df_{G0}, with an image $\hat{Y}_0$.

The dependent variable space H consists of N independently sampled functions y_i, each function being representable within a function space of dimension, say, K_H. If we use centered functions to fit model (3) the dimensionality of H is $df_H = (N-1)K_H$; otherwise it is NK_H. Dimensionality K_H in turn is the minimum of the number of basis functions used to represent each y_i and the number of sampling points at which each function is observed in the raw data. However, if a roughness penalty or regularization procedure is employed, assessing K_H is more complex, and a discussion of this issue can be found in Hastie and Tibshirani (1990) and Ramsay and Silverman (1997).

If the linear operator L is of full rank, then the dimension of the subspace of H containing the image $\hat{Y}_1 = L\beta_1$ for $\beta_1 \in G_1$ is df_{G1}; if the reduced model $\beta_0 \in G_0$ is involved, the dimensionality of the subspace containing $\hat{Y}_0$ is df_{G0}. It can be shown in this very general context from the nature of the inner product that, if the full and reduced models are fitted by minimizing $\|Y-\hat{Y}\|^2$, then H is partitioned into the direct sum of three orthogonal spaces containing $Y - \hat{Y}_1$, $\hat{Y}_1 - \hat{Y}_0$ and $\hat{Y}_0$, respectively; and that the dimensions of these subspaces are $df_H - df_{G1}, df_{G1} - df_{G0}$ and df_{G0}, respectively.

The procedure described above for estimating β will in general be of full rank, although in any actual application the design matrix $\mathbf{\Psi}$ should certainly be checked for nonsingularity. The squared norm is the squared error criterion

$$\|Y - \hat{Y}\|^2 = SSE = \int_0^T \sum_{i=1}^{N} \{y_i(t) - \hat{y}_i(t)\}^2 \, dt \ . \tag{9}$$

However, our fitting method is not precisely least squares because we actually minimize the discretized criterion $\|\mathbf{Y} - \mathbf{\Psi B}\|^2$. But, provided that the number of sampling points Q is sufficiently large, it seems appropriate to assume the orthogonal decomposition described above, and that the associated degrees of freedom are approximately correct.

Consequently we may be guided in model selection by

$$R^2 = 1 - SSE_1/SSE_0 \ , \tag{10}$$

and, moreover,

$$F = \frac{(SSE_0 - SSE_1)/(df_{G1} - df_{G0})}{SSE_1/(df_H - df_{G1})} \tag{11}$$

can be inspected as a means of assessing the importance of the global improvement in fit in going from the simpler model to the more powerful alternative. This use of the F-ratio as a model selection tool is based on the orthogonality of the subspaces of H containing $Y - \hat{Y}_1$ and $\hat{Y}_1 - \hat{Y}_0$, and that, when the reduced model holds, the lack of fit per degree of freedom in the two associated subspaces of H should be approximately equal. One should

probably stop short, however, of taking tabled values of the F-distribution as descriptions of its reduced model behavior since the numerator and denominator may not be assumed to have normalized chi-square distributions.

However, this type of global analysis may be of only limited interest if it seems clear that the improvement in fit depends strongly on t. Consequently, in addition to these two scalar measures of fit, we will want to display measures of fit as functions of t, for instance,

$$RMSE(t) = \sqrt{SSE(t)/N} \ , \tag{12}$$

where $SSE(t) = \sum_{i=1}^{N}\{y_i(t) - \hat{y}_i(t)\}^2$, which allows for closer time related analyses.

7 The regression of lip acceleration on EMG

A first issue is how fine to make the triangulation, which controls the smoothness and the amount of detail in the estimated $\beta(s,t)$. We used a preliminary calibration step in which we tested the method on a data set created from our actual EMG functions $x_i(s)$ using these intercept and regression functions:

$$\begin{aligned} \alpha(t) &= 2\sin(40t) + 2\cos(20t) \\ \beta(s,t) &= 0.5\sin(30s) + 0.5\cos(30t) \end{aligned} \tag{13}$$

For the model involving integration on the whole triangular domain ($\delta = M\lambda = T$), we computed the true dependent variable functions $y_i(t)$ with

$\epsilon_i(t) = 0$. We used triangulations into n^2 elements, $n = 4, 6, \ldots, 26$, where n is the number of intervals into which the horizontal and vertical boundaries of the triangular domain are divided. This means from 15 to 378 nodes, and from 16 to 676 triangular elements.

Using squared correlations R^2 defined in (10), we found, as expected, that the finer the mesh the better the fit. Figure 4 shows $\widetilde{\beta}(s,t)$ for 190 nodes and 324 elements, for which $R^2 = 0.997$. These results reassured us that the variation among the 32 EMG records was, in principle, sufficient to support estimation of a regression function with considerable local detail. For the analysis of our actual data, we chose a triangulation of the whole domain into 196 elements or 120 nodes, and this gave $R^2 = 0.991$ for this calibration problem. This triangulation corresponds to dividing $[0, T]$ into $M = 14$ intervals with lengths $\lambda = 50$ msec, roughly equal to the delay mentioned above necessary for a neural signal to be transduced into muscle contraction.

The second issue is the width of the domain of integration δ, which in this case is set at the discrete values $m\lambda$. From Figure 5, we see that the fit steadily improves as we enlarge the domain of integration from $\delta = 0$ ($R^2 = 0.17$) up to $\delta = 6\lambda$ ($R^2 = 0.46$), corresponding to 300 msec, but does not increase substantially with larger values.

Figure 6 shows the estimate of the bivariate regression function $\beta(s,t)$,

corresponding to $\delta = 6\lambda$. We estimated the standard error of $\beta(s,t)$ by drawing 1000 bootstrap samples from among the 32 acceleration functions. Figure 7 shows estimated β's values, with 95% pointwise confidence bands, for $\beta(s,t)$ along the diagonal $s = t$.

The shape of the estimate of $\beta(s,t)$ indicates, as expected, that the muscle activation is the most influential around the lip closure times, when the signal-to-noise ratio is the largest. Also, there is a ridge of influence along the diagonal, corresponding to the transduction delay of about 50 msec. But peaks deeper in the triangular domain reveal covariation of EMG events and lip displacements separated by longer delays. For instance, the peak at about $s = 250$ msec and $t = 500$ msec cannot be interpreted in terms of immediate causal relationship between muscle activation and lip acceleration. Rather it can be explained as follows. The agonist depressing activity of the DLI ends at around 250 msec, and this determines the position reached by the lower lip for the production of the /o/. The amplitude of the upward movement that follows depends on this position; and, consequently, so does the deceleration during the second part of this movement that is necessary to allow for an adequate closing of the lips for the second /b/. This illustrates how the pronunciation of a phoneme is determined by the chain of phonemes within which it is embedded.

Next, we specifically examined how well this version of the model performs

in comparison to the one using lag $\delta = 0$, that is, reduced to the pointwise model (2). Because there is inevitably less fitting power for values of $t < \delta$ than for t where the full history of EMG is available, we chose to consider only values of t going from 6λ to T.

R^2 values computed over this range of t are 0.57 and 0.22 for lags $\delta = 6\lambda$ and $\delta = 0$, respectively. Plotting the corresponding error functions $RMSE(t)$, defined in (12), in the top panel of Figure 8 shows that the model using the wider lag is superior specifically around the second closure time; making clear that the explanatory value of the past behavior of the EMG curves varies with the events to be explained. The small failure of the $RMSE(t)$ for $\delta = 6\lambda$ to be greater than that for $\delta = 0$ at $t = 0.36$ is due to the use of discrete values of t to estimate the models.

The bottom panel of Figure 8 shows F-ratio values integrated over each λ-wide intervals. Here, 100 basis functions were used to represent each of the $N = 32$ acceleration functions from data sampled at $n = 501$ points without regularization, so we have $K_H = 100$ and $df_H = 3100$. With $M = 14$, we have $df_{G1} = 84$ and $df_{G0} = 15$, so the degrees of freedom are 69 and 3016 for the numerator and the denominator respectively, and the critical value is $F_{.05} = 1.27$.

8 Discussion and Conclusions

Our application of the historical linear model (1) described the contribution of a facial muscle to the articulation of a phoneme. We could see that, in addition to the straightforward causal effects involving conversion of electrochemical into mechanical energy, more complex mediate influences take place over delays as long as 300 msec.

At least two groups of muscles applying force in different directions are generally necessary to control the motion of a body part, and three groups of muscles contribute to the motion of the lower lip. Our model, by using the activation pattern of only one single group of muscles, could explain only a limited part of the variation observed in the lip accelerations, and this contributed to the modest R^2 values that we obtained. Including the influences of the other two groups of muscles would have introduced an additional integral for each set of explanatory functions. However, we judged that the number of replications available could only support a single functional independent variable.

As with any linear model, there should be some careful consideration of whether the variation in the covariate $x(s)$ is sufficient to support a reasonably detailed fit to the observed dependent variable $y(t)$. In the multivariate context, this would correspond to checking the design matrix for near singularities. We felt that a calibration experiment such as the one that we used,

based on the actual covariates, offers a useful indication of whether a known regression function $\beta(s,t)$ can be recovered, and we found our results to be encouraging.

It is possible to envisage a number of approaches to the representation of the regression function $\beta(s,t)$, and to the estimation of the parameters or coefficients defining this representation. We chose the finite element approach because it conforms naturally to the triangular domain for $\beta(s,t)$, is easy to evaluate, permits an arbitrary amount of detail in the estimate, and leads to a sparse matrix linear equation defining the estimate. These are important assets when the estimate needs to have complex features. The finite element method has already proven itself in other areas of applied mathematics, and there are well-developed software tools available in languages like Matlab for tasks such as the triangulation of the domain (MathWorks, 1995).

A next step is to add the possibility of regularizing or smoothing the estimate of $\beta(s,t)$ or the prediction of $y_i(t)$ through the use of roughness penalties. These roughness penalties will require the use of basis functions $\phi_k(s,t)$ of higher order, but again these have already been well worked out in the finite element analysis literature.

Appendix

The raw data consisted of two-dimensional positions measured in meters in the sagittal plane sampled 625 at Hz. Jaw position was also recorded, and subtracted from lip position. Although two-dimensional position measurements were taken, in fact the trajectory of the lower lip was nearly linear, and consequently the data were reduced to one-dimensional coordinates by principal components analysis. The lip data were first smoothed by the robust lowess smoother in S-PLUS to eliminate the occasional outlying recordings. These data were in turn approximated using 100 B-spline basis functions. The spline basis was of order 6 in order to assure that the second derivative of the expansion would be reasonably smooth. A light roughness penalty on the fourth derivative was applied to the expansion in order to further smooth the second derivative, using the method described in Ramsay and Silverman (1997).

The EMG activity of depressor labii inferior (DLI) was recorded, in millivolts, using bipolar surface electrodes. Using Matlab software, EMG signals were analogue low-pass filtered at 600 Hz and digitally sampled at 1,250 Hz. The resulting signals were digitally band-pass filtered between 30 and 300 Hz, and full-wave rectified, i.e., absolute values were computed. To obtain the linear envelopes, the full-wave rectified signals were digitally low-pass filtered at 20 Hz, with a no-lag second-order Butterworth filter. These latter

were finally linearly interpolated to fit the lip data sampling rate.

Acknowledgments

We thank Prof. V. Gracco, McGill University, for making the data available. The research was supported by a grant to the second and third authors from the Natural Sciences and Engineering Research Council of Canada, and is based in part on the Masters Thesis by Malfait (1999).

References

AMES, W. F. (1992) *Numerical Methods for Partial Differential Equations.* third edition, San Diego: Academic Press.

BRENNER, L. R. & SCOTT, S. C. (1994) *The Mathematical Theory of Finite Element Methods.* New York: Springer.

HASTIE, T. & TIBSHIRANI, R. (1993) Varying-coefficient models. *Journal of the Royal Statistical Society, Series B.*, **55**, 757-796.

LENNEBERG, E. H. (1984) *Biological Foundations of Language.* New York: Wiley.

MALFAIT, N. (1999) Estimation par la méthode des éléments finis dans un modèle linéaire fonctionnel. Université du Québec à Montréal: Unpublished Masters Thesis.

MATHWORKS, INC. (1995) *Partial Differential Equation Toolbox User's Guide.* Natick, Mass.: The MathWorks, Inc.

RAMSAY, J. O. & SILVERMAN, B. W. (1997) *Functional Data Analysis.* New York: Springer.

Figure 1: The upper two panels show the position and acceleration of the center of the lower lip, respectively, during the articulation of the syllable "bob". The lower panel displays the linear envelopes of the EMG activity associated with the depressor labii inferior (DLI) muscle. The vertical dotted lines separate distinct segments in the motion of the lower lip.

Figure 2: The large triangle is the domain for the regression function $\beta(s,t)$ in the historical linear model. The horizontal axis s is time for the independent variable $x(s)$, and the vertical axis t is time for the dependent variable $y(t)$. The smaller triangles within the larger triangle, labelled $T^1, \ldots, T^{16}$, are finite elements used to approximate $\beta(s,t)$. The fifteen vertices of these triangles, called *nodes*, correspond to basis functions $\phi_k(s,t)$. The horizontal dashed line shows which elements play a role in determining the behavior of $y(t)$ at $t = 0.43$.

Figure 3: The basis function $\phi_{13}(s,t)$ corresponding to the 13th node in the triangulation shown in Figure 2.

Figure 4: The estimate of the regression function $\beta(s,t)$ for the calibration problem (13), corresponding to $R^2 = 0.997$.

Figure 5: The squared correlation measure (10) as a function of lag $\delta = m\lambda$.

Figure 6: The estimated regression function $\beta(s,t)$ for lag $\delta = 6\lambda$ for the actual data.

Figure 7: The solid line is the value of the estimated regression function along the diagonal, $\beta(t,t)$. The dashed lines indicate plus and minus two standard errors estimated by bootstrapping.

Figure 8: The top panel shows the error function $RMSE(t)$ for the models with lag $\delta = 6\lambda$ (solid line) and $\delta = 0$ (dashed line). The bottom panel shows F-ratio values integrated over each λ-wide intervals. In both panels, the vertical dotted lines indicate boundaries between elements.

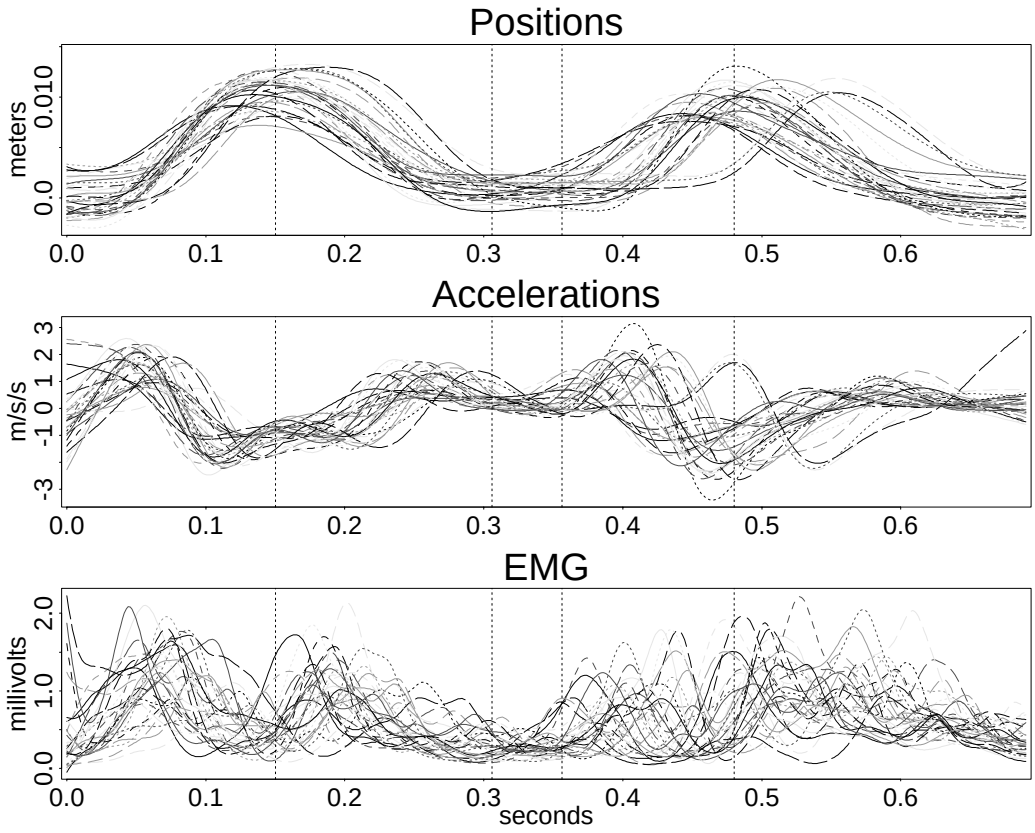
Positions
meters
0.0
0.010
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Accelerations
m/s/s
-3
-1
0
1
2
3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
EMG
millivolts
0.0
1.0
2.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
seconds

Figure 1:

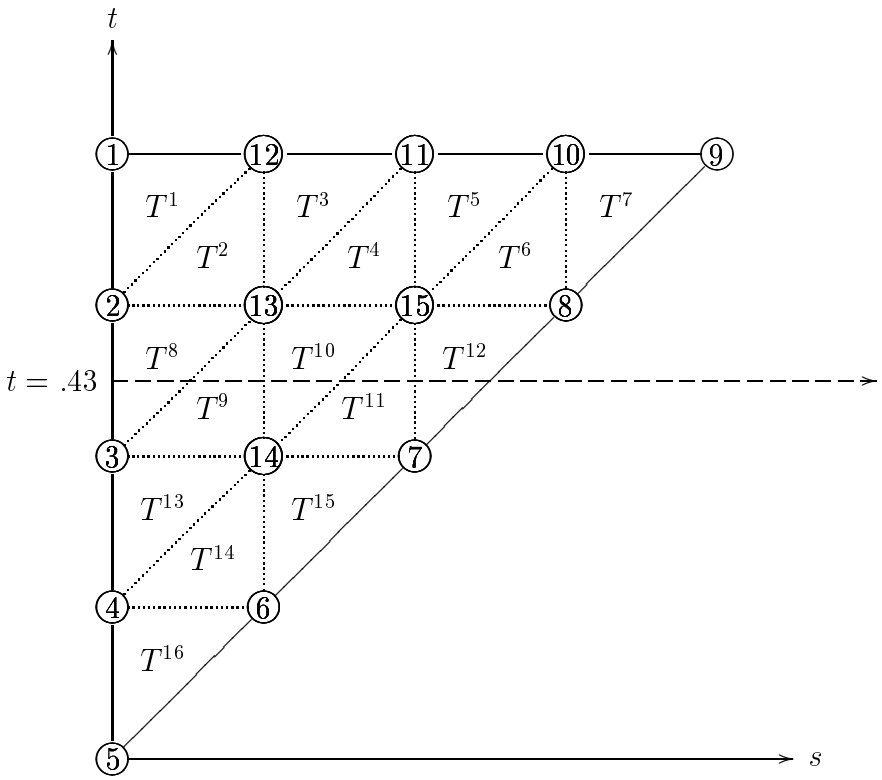
t
1
12
11
10
9
T1
T3
T5
T7
T2
T4
T6
2
13
15
8
T8
T10
T12
t = .43
T9
T11
3
14
7
T13
T15
T14
4
6
T16
5
s

Figure 2:

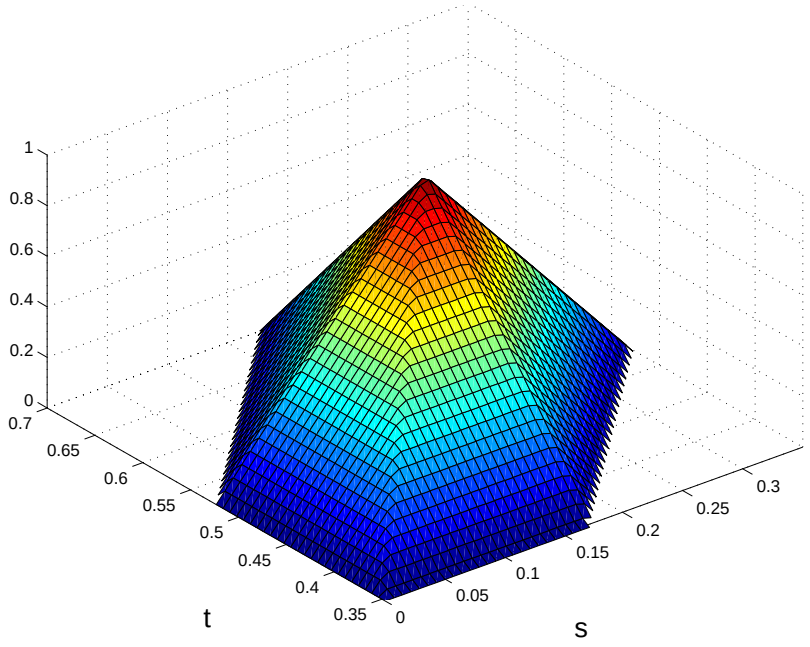

Figure 3:

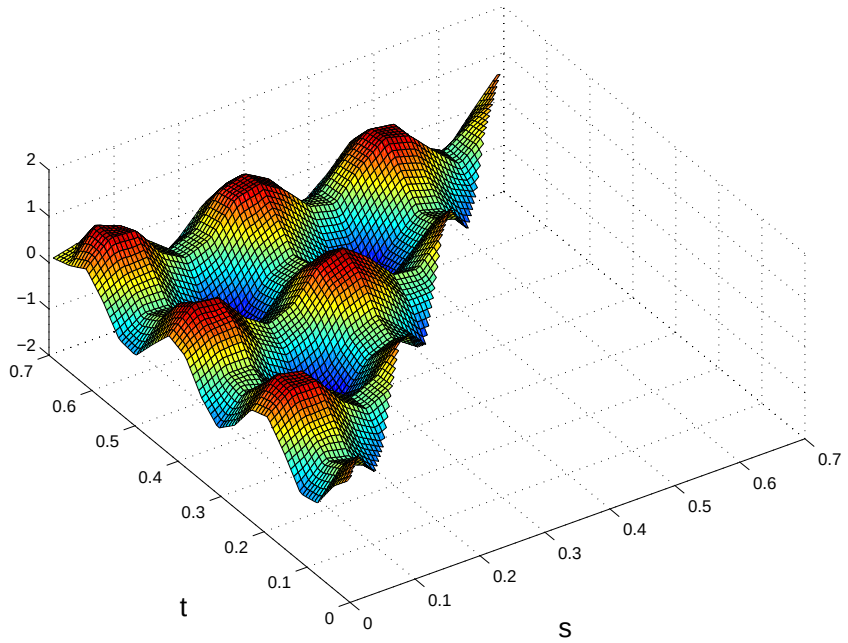

Figure 4:

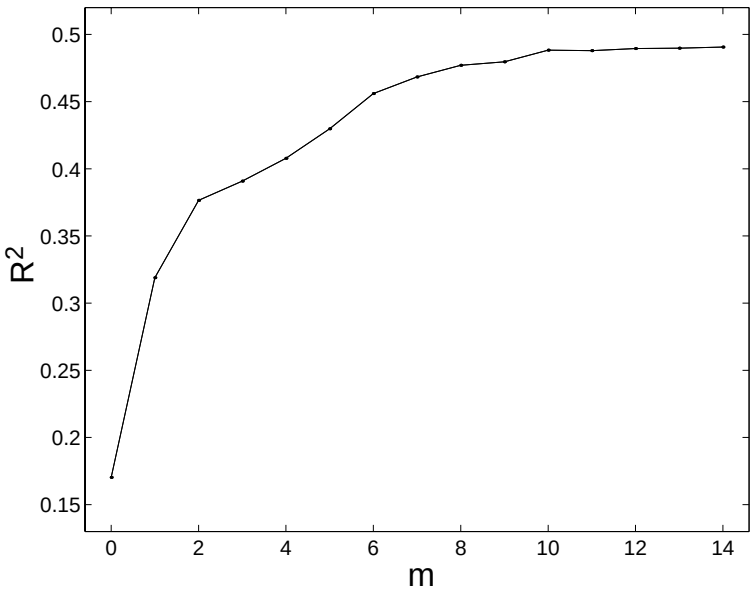

Figure 5:

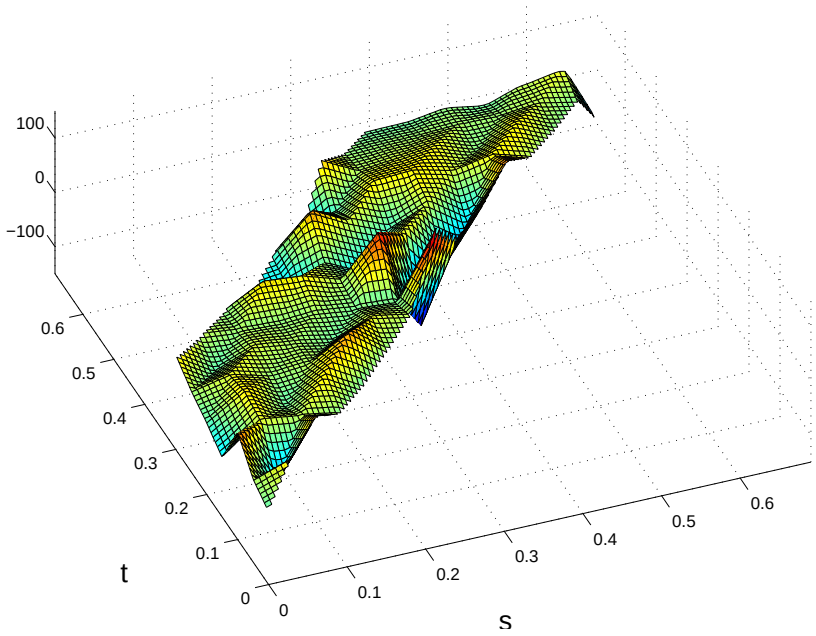

Figure 6:

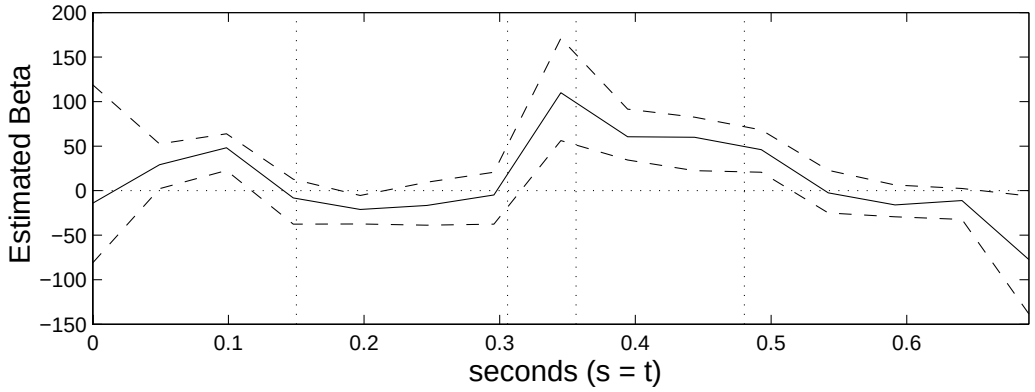

Figure 7:

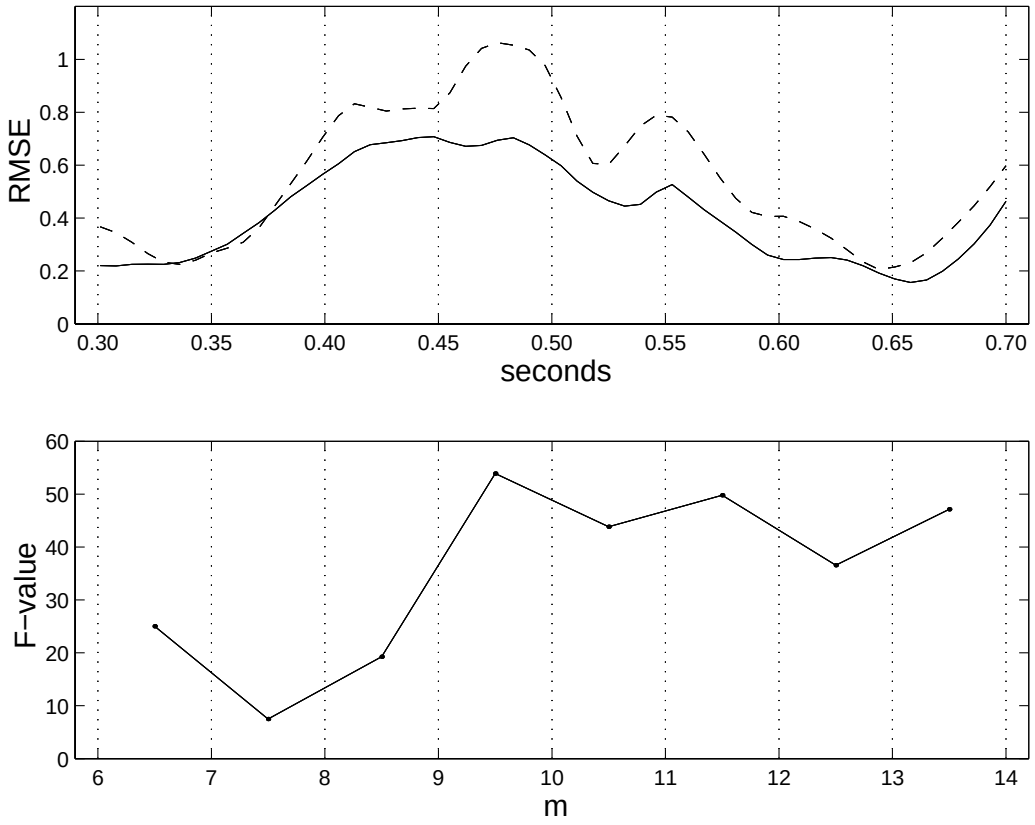

Figure 8: