

10

Predicting Lip Acceleration from Electromyography

10.1 The neural control of speech

Physiologists and psychologists who study motor control aim to understand how the brain controls movement. We know that waves of neural activation cascade down complex neural pathways to the motoneurons that activate muscle tissue, and that the contraction of these muscles applies forces to limbs. We know, too, from elementary physics that force is proportional to acceleration, and that if we study the acceleration of some body part, we are getting close to seeing how this remarkable control mechanism produces the movement that we see and feel.

Our capacity for speech is remarkable. In conversation, we can easily pronounce 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.

Does the brain plan sequences of speech movements as a group, or does it just control each movement in turn without regard to preceding or following phonemes? In speech production, the concept of *coarticulation* implies that the characteristics of each phoneme are adjusted to accommodate aspects of what is coming up ahead.

We can gain some insight into coarticulation by studying the lower lip. The lower lip plays a modest role in speech articulation, but is easily acces-

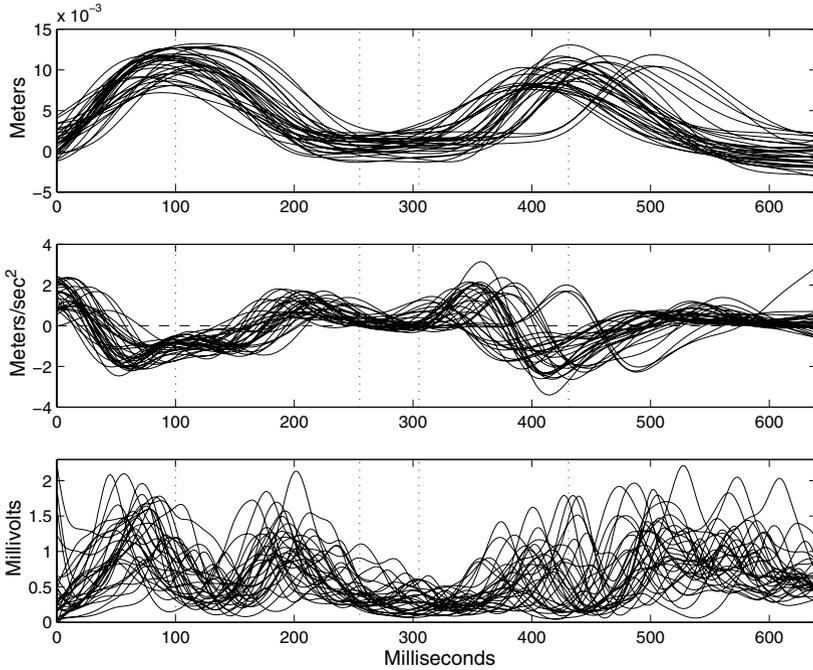


Figure 10.1. The top panel displays the position of the center of the lower lip of a speaker pronouncing the syllable “bob” for 32 replications. The middle panel displays the corresponding accelerations. The bottom panel contains electromyogram (EMG) recordings from a facial muscle that depresses the lower lip, the depressor labii inferior. The dotted lines indicate distinct phases in the articulation of the syllable. The EMG recordings are shifted to the right by 50 milliseconds, the time lag of the direct effect of a neural excitation as a muscle contraction.

sible, and is controlled by only three muscles. We can investigate how these muscles work together to control the lip, and how their contractions are determined by neural activation. We focus on the most important of the three, the depressor labii inferior (DLI) muscle that depresses the lower lip. To produce each /b/, the lip moves up to close the mouth, and then down. During these movements the DLI muscle plays specific roles: one, referred to as *agonist*, when it accelerates the lip during the descending phases, and the other, called *antagonist*, when it brakes the movement during the ascending phases.

Implanting electrodes to observe neural activity directly would involve more heroism than most subjects would consider worthwhile, but we can measure a byproduct of this activity through electromyographical (EMG) recording. Recordings are taken from the surface of the skin, and do not seriously perturb normal movement.

However, there are some issues with EMG recordings as indicators of neural activity. A muscle that is stretched in the absence of neural activation will also generate an EMG signal. Where muscles are overlapping or even just close together, the recording may not cleanly separate activity in different muscles. Finally, there is a period of about 50 msec following the onset of neural excitation, and the associated EMG signal, before muscle contraction begins.

Even if there is some imprecision in whatever EMG reflects, it cannot exert an influence backward in time on lip acceleration, since neural activity shows up in EMG signals with essentially no delay. Consequently we are interested in a *feedforward* model for the influence of EMG on lip acceleration. However, because of the 50 msec lag between neural activation onset and muscle contraction, only associations at delays substantially larger than 50 msec are evidence for coarticulation effects.

10.2 The lip and EMG curves

A subject was repeatedly required to say the syllable “bob,” embedded in the phrase, “Say bob again.” Because of the delay in muscle contraction indicated above, the records have been shifted in time, dropping the first 50 msec from the observed lip acceleration curves, and the last 50 msec from the raw EMG records. The duration of “bob” in each original record was time-normalized to 690 msec, but because of this time shift, only 640 msec is displayed in Figure 10.1.

The top panel of Figure 10.1 shows a sample of $N = 32$ trajectories of the lower lip. In the middle panel, the acceleration functions $Y_i(t)$ estimated from these original observations are shown. The bottom panel of Figure 10.1 shows the EMG records. The value $Z_i(s)$ plotted at any particular time s is the recording actually made at time s msec, but the values of lip position and lip acceleration plotted for the same time are those actually observed at time $s + 50$ msec, when any muscle contraction associated with activation at s msec is beginning to take place. Thus, for example, the last EMG observation plotted is for 640 msec, but the actual time for the corresponding lip observations is really 690 msec. For simplicity, we specify lip times from here on as the actual time less 50 msec. However, it is sometimes important to consider the real time of the observations, as we see below.

The lower lip trajectory can be segmented roughly into these epochs, separated by dotted lines in Figure 10.1:

1. close mouth for the first /b/;
2. lower the lip after utterance of first /b/;
3. central part of /o/, lip relatively stationary;

4. raise the lip for second /b/; and
5. lower the lip after the second /b/.

As we noted above, we can expect substantial EMG activity whenever the DLI muscle is active, whether the lip is descending or ascending. The point of least EMG activity is at about 330 milliseconds, at the end of the period when the lip is at its lowest point during the utterance of /o/.

How is the variability across observations of the EMG recording $Z(s)$ reflected in the behavior of the lip acceleration $Y(t)$? It is implausible to suppose that $Z(s)$ acts backward in time to influence $Y(t)$. Examination of Figure 10.1 may suggest that there is some forward influence of EMG activity on lip acceleration, but there is clearly statistical work to be done in investigating this possibility.

As a first step in studying the possible forward influence of EMG activity, we look at the correlation over the 32 replications of the electromyogram at times s and the acceleration at times $t \geq s$. The results are plotted in Figure 10.2. The light and dark patches on or very close to the diagonal of the image indicate a substantial amount of simultaneous relationship of both positive and negative polarity.

We can check for feedforward influence by scanning horizontally, to the left of the diagonal, for a fixed time t . For example, the dashed lines in the figure correspond to about 425 msec, when the lip is closed for the second /b/. We see a patch of positive correlation at about 350 msec; EMG activity 50 msec before this time, during the full opening of the mouth, is correlated with later acceleration. Further back, however, we see some negative correlation at about 175 msec, corresponding to EMG activity during the closure for the first /b/. As we scan parallel to the diagonal, we see a slightly curved band of positive correlation at a lag somewhere around 150 msec, and another band, but of negative correlation, further back at around 200 msec.

10.3 The linear model for the data

Let $T = 640$ indicate the final time of the complete utterance, and let δ be the time lag beyond which we conjecture that there is no influence of EMG activity $Z(s)$ on the lip acceleration $Y(t)$. With this in mind, we model $Z(s)$ as influencing $Y(t)$ according to the model:

$$Y_i(t) = \alpha(t) + \int_{t-\delta}^t Z_i(s)\beta(s, t) ds + \epsilon_i(t) . \quad (10.1)$$

Here $\alpha(t)$ is a fixed *intercept* function that allows for the relationship between the mean lip and EMG curves, but cannot accommodate their covariation effects. The model presumes that EMG affects lip acceleration

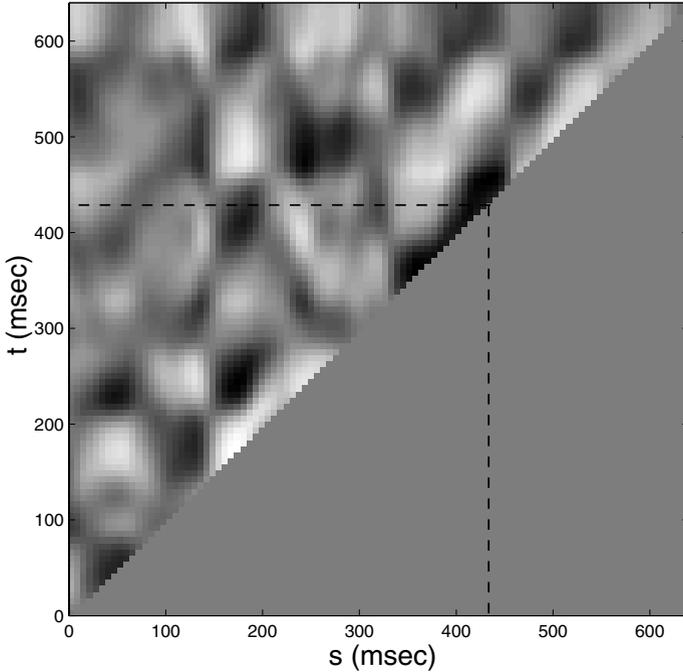


Figure 10.2. The correlations between the accelerations, functions of t , and the electromyogram recordings, functions of s , for all pairs of time values $s \leq t$. White regions correspond to positive correlations and dark regions to negative correlations. The gray level below the diagonal corresponds to a value of zero.

in a linear fashion, and the residual function $\epsilon_i(t)$ reflects the inability of the linear prediction model to fit the data completely. We might call this the *historical linear model* in the sense that the influence of $Z(s)$ feeds forward in time for a time lag of up to δ , and therefore is a relevant part of the history of $Y(t)$ for $s \leq t \leq s + \delta$. Since $s \leq t$, the regression coefficient function $\beta(s, t)$ is defined on a subset of the triangular domain used in Figure 10.2.

By contrast, the pointwise model

$$Y_i(t) = \alpha(t) + Z_i(t)\beta(t) + \epsilon_i(t) , \quad (10.2)$$

could be called *contemporary*, because the influence of EMG on lip acceleration is only instantaneous. In the contemporary model the regression function $\beta(t)$ depends only on t . The model can be viewed as a limiting version of the historical model as $\delta \rightarrow 0$.

The central question is, then, whether the contemporary model provides an adequate fit, or whether we should use a model in which β depends on both s and t . If we do fit a historical linear model, then we would also hope to gain some insight into the appropriate value of the lag δ .

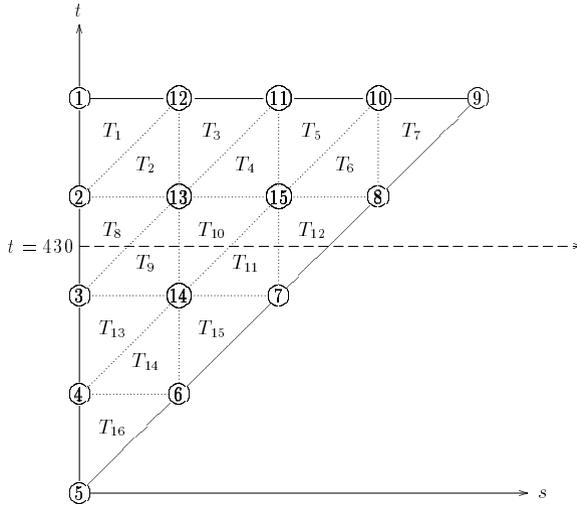


Figure 10.3. The domain of definition of the regression function $\beta(s, t)$ discretized into 16 triangular elements. Element boundaries are indicated by dotted lines and nodes by circled numbers. As an illustration, the horizontal dashed line at $t = 430$ represents the line of integration for $Y_7(430)$.

10.4 The estimated regression function

A practical approach to the estimation of the regression function $\beta(s, t)$ is to seek an expansion in terms of a fixed number of known basis functions. We use the *finite element method*, often used in engineering to solve partial differential equation systems. This approach involves subdividing the domain (s, t) , $s \leq t$, into triangular regions in the manner shown in Figure 10.3. The triangles are called the *elements* and the vertices of the triangles are called the *nodes* of the system. Sixteen triangles are shown in the figure, corresponding to four intervals along each axis; but our final triangulation involved 169 elements and 105 nodes, resulting from using 13 intervals along each axis, each interval being of length $640/13 = 49.2$ msec.

The next step is to define basis functions over each of these regions. Each basis function $\phi_k(s, t)$ is a linear bivariate function having the value one at a specific node and falling off to zero at the remote edges of each triangle that has that node as a vertex. A typical basis function for a node inside the triangular domain is shown in Figure 10.4.

The triangular basis has an important advantage in considering how large the lag δ should be in modeling the feedforward influence of $Z(s)$. Triangles falling more than 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 reasonably fine. Letting Δ indicate the width

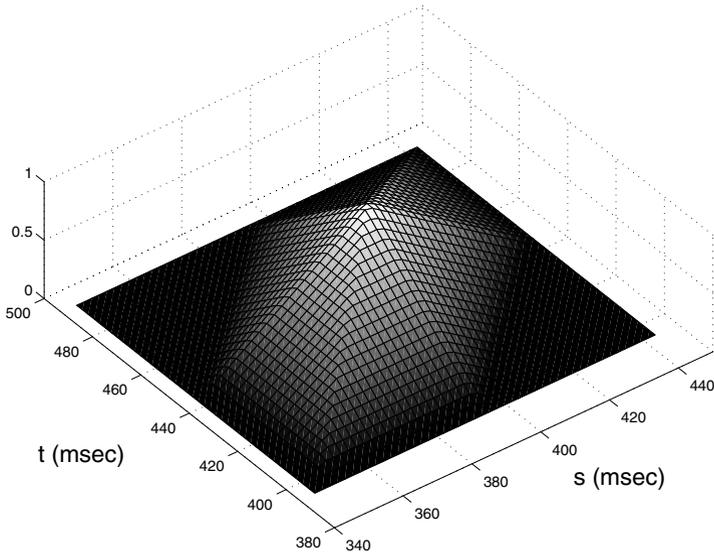


Figure 10.4. A typical piecewise linear basis function used to construct a finite element approximation of the regression function $\beta(s, t)$.

of a single triangle, we are permitted to use lag values $\delta = m\Delta$ for integers $m \geq 1$.

The contemporary model (10.2) can be thought of as the case $m = 0$. In this case, the elements are intervals along the diagonal line. The basis functions are functions of only one variable t , and are piecewise linear functions, in other words, B-splines of order 2, as shown in Figure 2.14.

Once we have estimated the coefficients b_k , we have a piecewise linear approximation to the regression function $\beta(s, t)$. The process of estimating the coefficients of the expansion can be set up as a matrix computation; for further details, see Malfait, Ramsay, and Froda (2001).

Figure 10.5 shows the full bivariate regression function $\beta(s, t)$, effectively setting $\delta = T$, as a grayscale image. For what values of t is lip acceleration most influenced by current and previous EMG activity? We see that the patterns of relationship that we already observed in Figure 10.2 are also found here, but the regression function surface is much better at picking out specific intervals where the influence is important. The peaks and dips in $\beta(s, t)$ indicate that the lip activity is most influenced by measured EMG in the time interval from about $t = 350$ to about $t = 480$, the time of the second lip closure. By scanning along the line corresponding to 425 msec, we note that there is also some indication that EMG activity at time $t = 250$, at the beginning of /o/, influences the second /b/ closure.

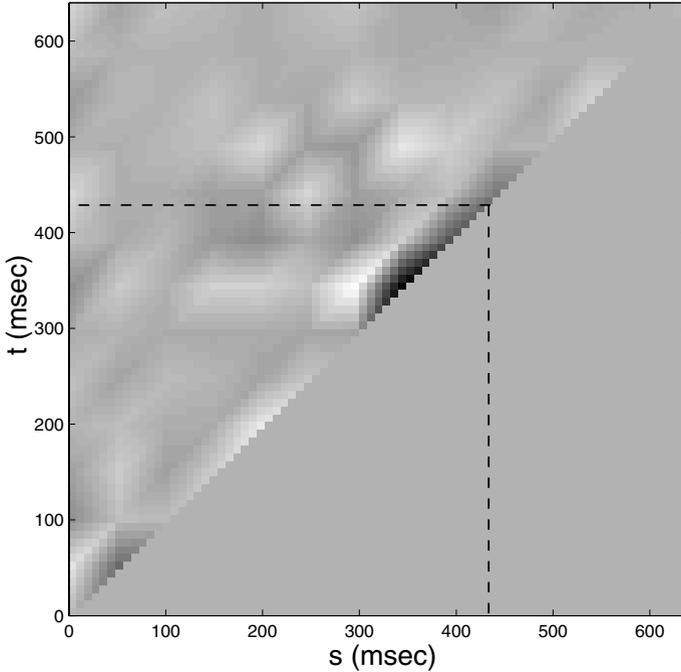


Figure 10.5. The regression function $\beta(s, t)$ estimated using 105 nodes. Dark regions correspond to negative values and white regions to positive values. The gray level plotted below the diagonal corresponds to the value zero.

10.5 How far back should the historical model go?

What lag δ seems to be supported by the data? To answer this, we need to compare a fit for a specific lag to that offered by a simpler, and more restricted, model. Two simpler models are the mean computed across the 32 replications,

$$\bar{Y}(t) = N^{-1} \sum_{i=1}^{32} Y_i(t),$$

and the contemporary model (10.2).

For a specific $\delta = m\Delta$, we can define the error sum of squares function at any time t by

$$\text{SSE}_m(t) = \sum_{i=1}^N \{Y_i(t) - \hat{Y}_i(t)\}^2, \quad (10.3)$$

where $\hat{Y}_i(t)$ is the fit of the current model to the observed curve $Y_i(t)$. For any given value of m , the squared multiple correlation measure of fit R_m^2

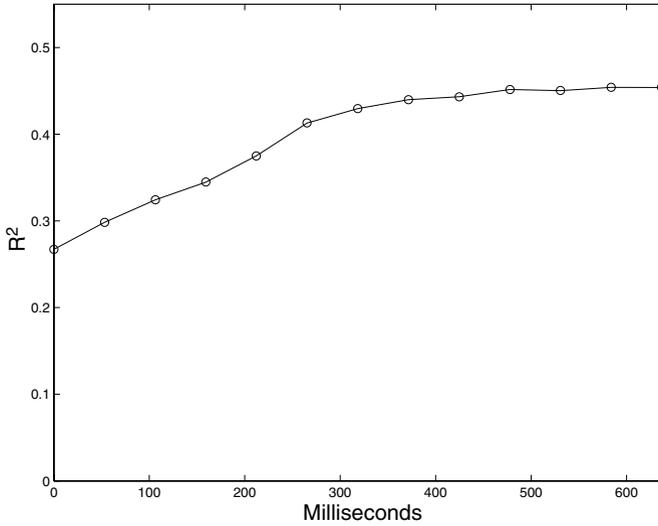


Figure 10.6. The squared correlation R_m^2 as a function of lag $\delta = m\Delta$ for a triangulation into 169 elements and 105 nodes. The points plotted correspond to the discrete values of δ given by integers m .

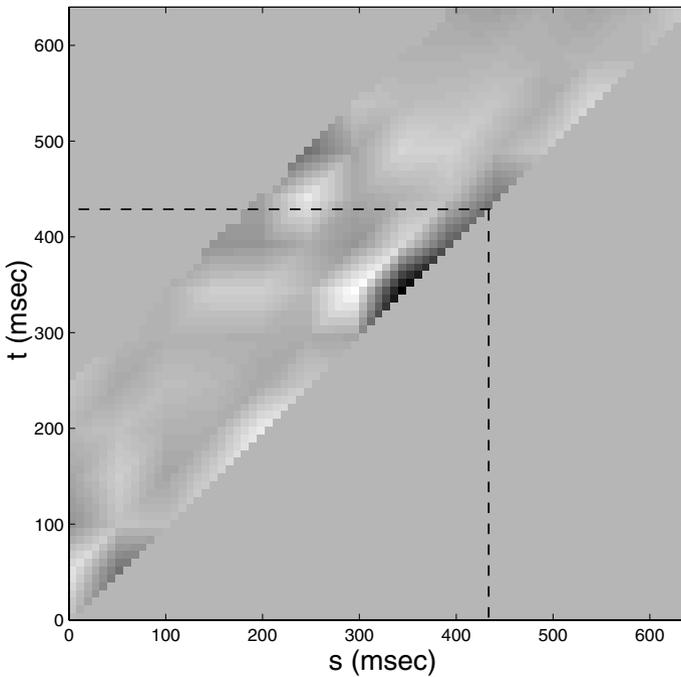


Figure 10.7. The estimated regression function $\beta(s, t)$ for lag $\delta = 5\Delta$.

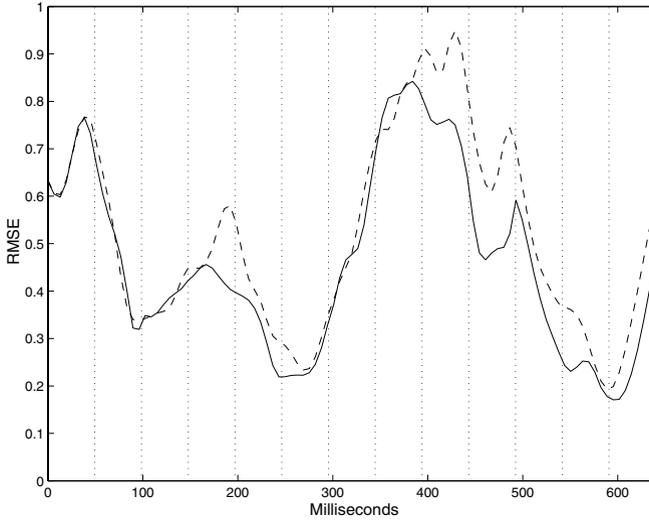


Figure 10.8. The error function $RMSE(t)$ for the models with lag $\delta = 5\Delta \approx 250$ (solid line) and $\delta = 0$ (dashed line). The vertical dotted lines indicate the positions on the axis of the nodes of the finite element basis.

is defined by

$$R_m^2 = 1 - \frac{\int_0^T SSE_m(t) dt}{\int_0^T SSY(t) dt}, \tag{10.4}$$

where $SSY(t)$ refers to the fit using the mean curve.

We consider R_m^2 as a function of $\delta = m\Delta$, that is, as a function of the width of the domain of integration in the model (10.1). From Figure 10.6, we see that the fit improves as we enlarge the domain of integration up to $\delta = 5\Delta$, but does not increase substantially with larger values of δ . Thus, it seems to be worth modeling lip acceleration at time t to be influenced by EMG values up to about 250 msec before t . The estimate of $\beta(s, t)$ using this lag is shown in Figure 10.7.

The shape of the estimate of $\beta(s, t)$ indicates, as expected from the regression function already considered, that the muscle activation is the most influential in the period leading up to the second lip closure times. Also, there is a ridge of influence along the diagonal continuing for a short period after the closure; in this short interval it is only contemporary EMG signals that matter. This suggests that the system plans the closure, but the recovery after the closure is not planned for in advance.

Figure 10.8 compares the standard deviation function $RMSE(t) = \sqrt{SSE_m(t)/N}$ for the historical model with $m = 5$ with that for the contemporary model $m = 0$. We see that the main improvement for the historical model is in the articulation of the second /b/ between 400 and 500 msec,

and also more briefly at about 200 msec, in the transition from the first /b/ to the /o/ phoneme.

Is the fit of the historical model with $m = 5$ significantly better than that of the contemporary model? Because different finite element bases are used to approximate the two models, the finite element contemporary model is not exactly nested within the finite element historical model, even though the exact models can be regarded as nested. In order to compare nested models, therefore, we approximate the contemporary model by the historical model with $m = 1$, and construct an F -test of significance. Results reported in full in Malfait, Ramsay, and Froda (2001) then demonstrate that the fit of the model with $m = 5$ is indeed significantly better than the approximate contemporary model $m = 1$.

10.6 What have we seen?

It now seems fairly clear from these results that the timing and intensity of phonemes do have a covariation with EMG activity that is reflected both in the simple correlation plot in Figure 10.2, and in the feedforward linear model (10.1). The time lag over which this feedforward influence is evident is not unlimited, and in this case corresponds to two phonemes. Of this 250 msec lag, we are able to account for 45% of the variation in $Y(t)$ by its covariation with $Z(s)$. This is a substantial effect, considering how volatile EMG data tend to be, as well as their tentative connection with neural activity. The pointwise or contemporary linear model only explains about 27% of the variability, and Figure 10.8 indicates that its deficiency as a model seems mainly concentrated on the second “b,” where the feedforward influence is especially strong.

Ramsay and Silverman (1997, Chapters 9 to 11) give a general introduction to functional linear models, and discuss various aspects in more detail. However, their treatment does not go as far as the restriction of the influence to a finite lag, and the present case study exemplifies the way that functional data analysis methods often have to be tailored to the particular problem under consideration. The finite element method adopted was particularly appropriate to the restriction to finite lag on the triangular domain over which $\beta(s, t)$ is defined. This approach also allowed a simple control of the size of the lag δ so that we could explore the role of this parameter.

10.7 Notes and bibliography

The data were collected at the Haskins Speech Laboratories at Yale University by V. Gracco. The analyses of the data were carried out by N.

Nicole during a Masters of Science program at the Université du Québec at Montréal, and reported in more detail in Malfait, Ramsay, and Froda (2001).

The raw lip position data consisted of two-dimensional positions in the sagittal plane sampled 625 times per second. 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.

A considerable amount of preliminary processing was required before satisfactory acceleration curves could be produced. The data were first smoothed by a robust method, the LOWESS smoother (Cleveland, 1979) in the S-PLUS package to eliminate the occasional outlying recording. These smoothed 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 in order to smooth the second derivative further.

The EMG data were sampled at 1250 hertz, and were much noisier than position records, showing very high frequency oscillations about zero as well as the slower trends that interest us. As is usual for EMG measures, the raw data were replaced by values of a linear envelope of the absolute values. These values were then further smoothed.

The contemporary model (10.2) can be viewed as a functional extension of the *varying coefficient model* of Hastie and Tibshirani (1993).