Functional Data Analysis in Action

J. O. Ramsay, McGill University
P. Gribble, Queens University
J. O. Ramsay, 1205 Dr. Penfield Ave., Montreal, Quebec, Canada H3A 1B1.

Key Words: Juggling, differential equations, dynamic model, registration, principal differential analysis, time warping

1 Introduction

These notes describe our analyses of data collected from Michael Newton, Department of Biostatistics, University of Wisconsin, in November 1998. We recording him juggling 3, 4, and 5 balls as well as bowling pins.

Our goal was to model the movement of his hands during juggling by a relatively simple second order differential equation. This study was a follow-up on our earlier work on handwriting (Ramsay, 1999), where a second order linear differential equation was able to account well for pen movement during the writing a complex piece of Chinese script. The larger aim of these studies is to shed light on the way in which the brain controls movement during complex task. A differential equation permits the study of the dynamics of this process, and fits simultaneously a number of derivatives as well as the position data. Since neural control, leading to muscle contraction, must, by the laws of mechanics, have a direct impact on acceleration, differential equations seems to be a natural modeling framework.

The analyses in this report use the techniques for functional data analysis described in Ramsay and Silverman (1997), and the software in Matlab and S-PLUS available at the Internet site `ego.psych.mcgill.ca/pub/ramsay/FDAfuns`.

2 The Data

The data were collected in four experimental conditions: Juggling three, four, and five balls, and juggling three bowling pins. In each condition, there were 10 trials, each lasting 10 seconds. During each trial, the three dimensional coordinates of infra-red emitting diodes (IREDs) were recorded 200 times per second by an OPTOTRAK system. The IREDs giving the results in this paper were positioned (i) on the index finger and the wrist at the base of the thumb of the right hand, and (ii) on the left, center, and right of the chest. In order to keep the report brief, will only consider the motion of the index finger while juggling three balls.

This research was supported by grant A320 from the Natural Sciences and Engineering Research Council of Canada. The authors thank D. Ostry, in whose laboratory the data were collected.

3 Preliminary Analyses

3.1 Smoothing the Data

The first step was to smooth the data so as to be able to fill in missing values, and consequently, to obtain estimates of all positions for each trial at 2000 time points, equally spaced from 5 msec to 10000 msec. Our smoothing tool was the Pspline smoothing package available for the SPLUS environment at `statlib.stat.cmu.edu`. This is a polynomial spline smoothing system that permits the specification of the order of derivative in the penalty term. We chose to penalize the derivative of order five for each coordinate, since we knew in advance that we would need smooth derivatives of order three for subsequent analyses.

If we indicate the value of coordinate k among the three spatial coordinates at time t by $P_k(t)$, its derivative of order m by $D^m P_k(t)$, and the observed coordinate value at time t_j by Y_{jk}, then our smooth is defined by minimizing the criterion

$$\texttt{PENSSE}_\lambda = \sum_j^{2000} [Y_{jk} - P_k(t_j)]^2 + \lambda \int_0^{10} [D^5 P_k]^2(t)\, dt \ . \tag{1}$$

We experimented with various values of the smoothing parameter λ, looking at values of the generalized cross validation (GCV) criterion, at the extent to which residuals had a white noise distribution, and at the smoothness of the higher order derivatives. We settled on a value of $\lambda = 10^{-12}$, which gave good fits to the more variable finger data, smooth but detailed 3rd derivatives, but tended to oversmooth slightly the less variable chest coordinates. The residuals for the finger IRED had a standard deviation of about 0.5 mm, and within the con-

text of a vertical variation of about 30 cm, this represents a signal-to-noise ratio of about 600 to 1.

We use the notation $V_k(t)$ to refer to the velocity function for coordinate k, $A_k(t)$ for its acceleration function, and $J_k(t)$ for its third derivative (sometimes called "jerk".) We will also make use of tangential velocity, defined by

$$TV(t) = \sqrt{V_1^2(t) + V_2^2(t) + V_3^2(t)} \ . \qquad (2)$$

This is the length of the velocity vector in 3-space, and measures the total magnitude of velocity at time t, and $TV^2(t)/2$ is the kinetic energy at time t. Tangential acceleration is correspondingly defined by

$$TA(t) = \sqrt{A_1^2(t) + A_2^2(t) + A_3^2(t)} \ . \qquad (3)$$

3.2 Centering and Rotating

The coordinates used by the OPTOTRAK system are arbitrary with respect to the geometry of the body and the juggling sequence. The next step was to define a coordinate system that would be more meaningful, as well as more appropriate for computation.

We observed some gentle and slow change of chest position and orientation over the 10 seconds of a given trial. These chest coordinates were low-pass filtered to eliminate frequencies higher than 1.5 herz, and we removed large scale upper body movements within a trial by averaging the three filtered chest coordinates for each time, and subtracting this average curve from finger and wrist coordinates.

The second step was to subtract the mean coordinate value taken across the 2000 values from each these values, so that each coordinate had a mean value of 0 within each trial.

Finally, we rotated so as to leave coordinate 3, the vertical coordinate, unchanged, but so that coordinate 1 exhibited the greatest variation in the horizontal plane. This first rotated coordinate corresponds closely to lateral movement across the body plane, with positive sense being from the midline to the right, from the perspective of the juggler. The second rotated coordinate reflected primarily fore-and-aft motion, within the saggital plane, and the positive sense being away from the body. These coordinates are, however, not themselves "natural" coordinates with respect to body dynamics, but merely convenient viewing coordinates from a spectator's perspective.

3.3 Trimming and Decomposing Records

Each 10 second sequence contained from 12 to 13 complete cycles of the hand throwing and then catching the ball. It was easy to recognize features in derivatives through order 3 that were stable across cycles and trials.

Because the timing of the beginning of a record is relatively arbitrary, partial cycles at the beginning and end of each record must be trimmed off to obtain records that are comparable across trials. The tangential acceleration for the finger IRED showed a strong and stable peak preceded by a deep minimum within each cycle, and the beginning of a cycle was defined by the location of this minimal value of $TA(t)$.

We then cut each record into cycles using the same features. This gave us 123 complete cycles. The average duration of the cycles was 711 msec, and 50% of their durations ranged from 696 to 736 msec, or over about 6% of the mean.

3.4 Registration

The registration process is the estimation of a transformation of time for each record so that cycle features will be more or less aligned. This is an essential preliminary to any cross-cycle analyses, including simply computing an average. In effect, the registration process separates *phase* or temporal inter-cycle variation from *amplitude* or spatial variation. These two types of variation can then be analyzed separately.

A first step was to linearly transform the duration of each cycle so that it extended over the mean duration of 711 msec. The typical cycle had about 142 values, and these were interpolated using the Pspline package to exactly 151 equally spaced values. It was necessary at this point to apply a higher level of smoothing to stabilize the behavior of the 2nd and 3rd derivatives near the beginning and ending of each record, where the lack of information about derivative behavior can lead to some wild variation in higher derivatives. To achieve this, and to convert the discrete observations to functional data observations, we turned at this point to expanding each cycle in terms of a basis defined by 55 order 6 b-splines with 51 equally spaced knots.

Following this linear registration, cycles could be considered to begin and end at a common point. But the timing of salient features inside the cycles could still be seen to vary considerably from cycle to cycle, and especially in the higher derivatives. Consider the first panel in Figure 1, for example. It shows the 123 tangential velocity curves superimposed prior to

registration. While can see some general features, these obviously vary in timing, and this variation obscures some of the finer detail.

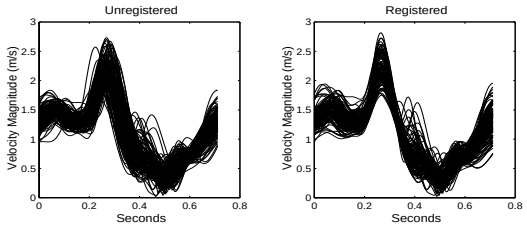

Figure 1: The left panel displays all 123 tangential velocity $TV_i(t)$ functions before registration, and the right panel shows registered functions.

What we wanted was a nonlinear time warping function $h_i(t)$ for cycle i with beginning and ending values 0 and 0.711 seconds, respectively, but with sufficient nonlinear behavior within the interval to align features for cycle i to those for a common target. Recall that each cycle is a 3-dimensional function with values $\mathbf{P}_i(t) = (P_{i1}(t), P_{i2}(t), P_{i3}(t))'$, so that $h_i(t)$ must align the three coordinates simultaneously.

We opted for the nonlinear registration process described in Ramsay and Li (1998), and implemented in Matlab in code available by ftp from the site given above. Based on results in Ramsay (1998), this method expresses the warping function as

$$h_i(t) = C_i \int_0^t \exp \int_0^u w_i(v)\, dv\, du, \qquad (4)$$

where constant C_i is defined by the requirement $h_i(0.711) = 0.711$ seconds. It is the function $w(t)$ that is actually estimated in this process, and this is the *relative curvature* $D^2h(t)/Dh(t)$ of the warping function. Function $w(t)$ was defined by using a basis consisting of 13 order 4 b-splines with equally spaced knots. Moreover, we smoothed the warping function by controlling the size of $w_i(t)$, since $w_i(t) = 0$ implies a linear warping function. Consequently, we attached to the fitting criterion the penalty $\gamma \int w^2(t)\, dt$, with $\gamma = 0.01$.

The registered functions $\mathbf{P}^*$ are set up as follows:

1. compute the inverse warping functions $w_i^{-1}(t)$ by smoothing the relationship between t itself as the dependent variable and $w_i(t)$ as the independent variable, and
2. then smooth the relationship between $\mathbf{P}_i(t)$ and $w_i^{-1}(t)$ to get $\mathbf{P}_i^*(t)$.

The common target used for registering the cycles is computed in two phases. First, the simple cross-sectional mean of unregistered tangential velocity was computed. Of course, because the $TV_i(t)$ are not registered at this point, the mean lacks a lot of the finer detail in the individual curves. But after registration, the mean $\bar{TV}(t)$ is recomputed since the partially registered curves now give a much cleaner picture of the mean. Using this target, the curves are registered again. While this process could be repeated, experience has convinced us that two iterations is sufficient.

The registered tangential velocities are given in the second panel of 1, and we now see much more alignment in curve features. The functions $h_i(t) - t$, called deformation functions, are shown in the first panel of Figure 4, and corresponding functions $w_i(t)$ in (4) in the second.

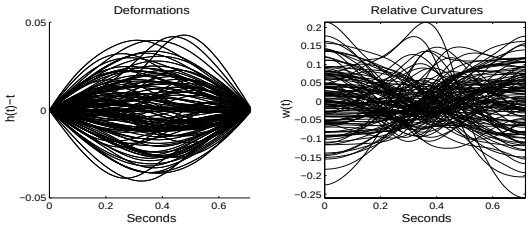

Figure 2: The left panel displays the deformation functions, $h_i(t) - t$, used to register the cycles, where the warping function $h_i(t)$ is defined in (4). The right panel shows the relative curvature functions $w_i(t)$; $w(t) = 0$ implies linearity.

4 Features in the Average Cycle

Figure 3 displays all three mean coordinates and their first two derivatives. Figure 4 shows the average position of the index finger in terms of coordinates X and Z, or vertical/lateral, Y-Z or saggital, and X-Y or horizontal. The overall shape of the cycle in the X-Y plane is that of an inverted pear, with something nearly like a cusp or point at the bottom, preceded by a small lateral movement. Progress around the loop is anti-clockwise, moving from the far right to the top, over and down the midline, pausing at the cusp at the bottom, and moving up and away from the midline.

In the horizontal plane, however, the effect is of a smooth figure 8, although with a sharp corner at the bottom corresponding to the cusp in the X-Z plane. Thus, while most of the variation is in coordinates X and Z, there is also a strong twist that implies that all three coordinates are essential.

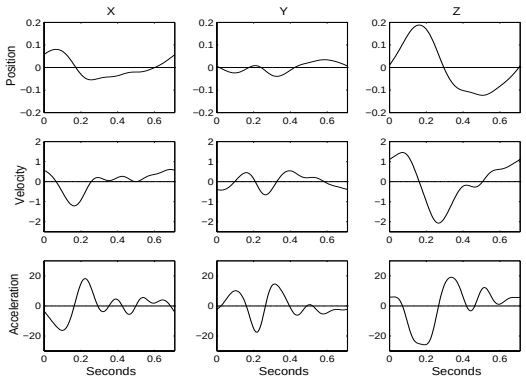

Figure 3: The mean position, velocity, and acceleration functions for each coordinate.

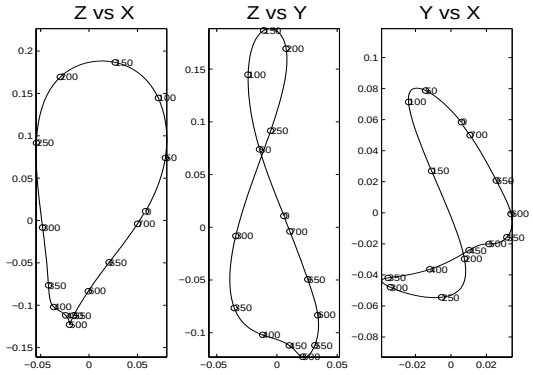

Figure 4: The left panel plots the vertical coordinate against the lateral coordinate (Z vs X), the center planel plot the vertical coordinate against the fore/aft coordinate (Z vs Y), and the right panel plots the lateral coordinate against the fore/aft coordinate (Y vs X). Numbers and small circles indicate time in milliseconds.

The cycle begins and ends in the middle of the upward movement following the throw of the ball. At that point the ired is about half way between the perigee and apogee. The apogee is reached at about 150 msec, and the maximum height is 19 cm. The arc is open and gentle at the top. The bottom of the cycle consists of two minimal heights. The first is at about -11 cm, and is reached at 390 msec. But at this point $\bar{P}_2(t)$ is still decreasing, while $\bar{P}_3(t)$ has bottomed out slightly earlier. So at this point the motion is more lateral than up and down. The perigee, reached at 510 msec, is at height -13 cm, so that the total height variation is 31 cm.

Additional information is available in Figure 5 that displays the average tangential velocity and acceleration. Using these, we see that the acceleration magnitude at the apogee is about $27m/s^2$, and the velocity at this point is fairly high, about $1.3m/s$. The ball is caught about about 210 msec, when we see a sharp peak in the tangential acceleration, with a value of $35m/s^2$. This acceleration pulse is the strongest in the cycle, as the ball imparts its energy to the finger. The ired has just passed the apogee, is beginning to decend, and has reached a height of 15 cm. Following the contact with the ball and the acceleration pulse, there is a sharp increase in tangential velocity, which reaches its maximum of $2.2m/s$ at 260 msec, as the hand descends rapidly towards the perigee. There is a third strong acceleration pulse with a peak of $24m/s^2$ at 330 msec that will decelerate the ball. At the cusp-like perigee, there is a sharp upward acceleration to achieve the throw, with the tangential value being $13m/s^2$. Just before the end of the cycle, at 670 msec, there is a hint of a small upward acceleration peak, possibly due to the ball leaving the hand.

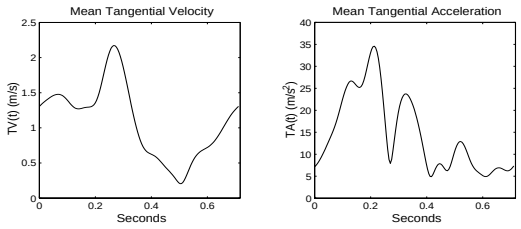

Figure 5: The left panel shows the mean tangential velocity, and the right panel the mean tangential acceleration.

5 Principal Components Results

The functional versions of principal components analysis in Ramsay and Silverman (1997) show the nature of the variability from cycle to cycle in aspects of the curves. In the registration phase, we separated this variability into phase and amplitude components.

5.1 Phase Variation

Beginning with the phase variation, the principal components analysis of the functions $w_i(t)$ defining the warping functions $h_i(t)$ indicated that two principal components account for 97% of the variation displayed in Figure 2. The first component reflects early versus late variation, and accounts for 74% of the variation, while the second component is early followed by late or its complement. The warping functions corresponding to these two components are displayed in Figure 6, The basis used to expand the $w_i(t)$'s was sufficiently powerful to reveal rather more complex modes of variation than these two, and seems remarkable that the phase variation is so simple. But perhaps this is because the hand must track the ball, whose time in the air is governed primarily by the strength of the vertical component of the throw by the other hand.

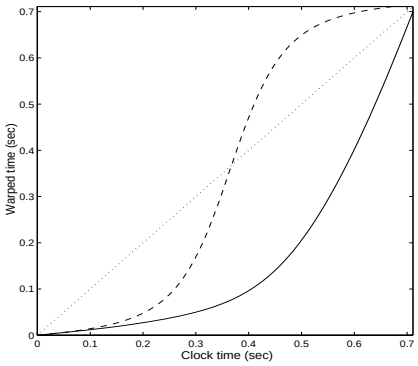

Figure 6: The warping functions $h(t)$ corresponding to the two dominant principal components of variation among the functions $w(t)$. The solid line component accounts for 73.5% of the variation, and the dashed line component for 23.8%.

5.2 Amplitude Variation

The registered cycles showed four dominant modes of variation, accounting for a total of 78% of the variation. The first 15 log eigenvalues are plotted in Figure 7. We find that in most functional PCA's, higher order log eigenvalues tend to decay linearly, and the four dominant eigenvalues are large relative to this linear decay.

We applied a VARIMAX rotation of the four principal component functions to aid in their interpretation. These appear to be as follows:

1. (33%) This is primarily in the fore/aft-direction. Temporally it begins a strong positive perturbation at the catch at 210 msec, crosses over the mean at around 42 msec, and ends its negative phase at the end of the cycle. It may be a compensation for a ball caught further away than normal from the body, and moved closer than usual to the body to maintain overall equilibrium.

2. (13%) This involves all three coordinates, and is confined to the beginning of the cycle until the catch. It may be, in fact, a continuation of the variation in the first component.

3. (15%) This seems similar to component 1 in shape, but also involves vertical motion. Perhaps it reflects a ball caught too high as well as too far from the body.

4. (14%) This is like component 1. but involves only lateral motion, and corresponds to a catch too far to one side.

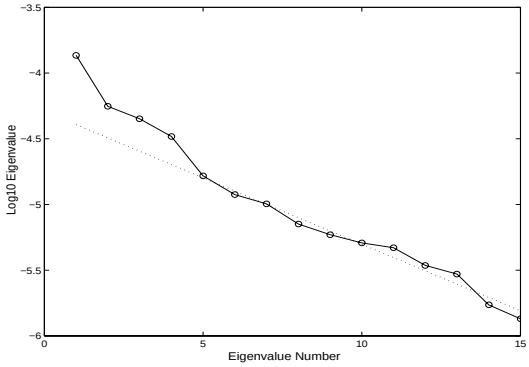

Figure 7: The logarithms of the eigenvalues associated with the principal components analysis of the coordinate variation.

These four types of variation appear, then, primarily to reflect the responses to variations in the direction of throw.

6 The Linear Differential Equation

As we indicated at the beginning of the paper, a central goal of this study for us was model these data by a linear ordinary differential equation. For the handwriting data reported in Ramsay (1999), a second order linear equation was able to do a good job of capturing both the shape of the mean curve, and a substantial part of the curve-to-curve variation. This equation modeled velocity rather than position so as to render it invariant with respect to translations of the coordinate system.

We tried the same approach in this context in the sense of estimating a second order equation in velocity separately for each coordinate. The results were encouraging, but the rationale for treating each coordinate separately seemed flawed. Whereas in the handwriting the coordinates that we used could be argued to be approximately "natural" given that handwriting tends to be organized into vertical and horizontal strokes and lifting the pen off the paper, here there is no such natural coordinate frame. And if there is, it seems unlikely that it does not change with time over the cycle, since the wrist and finger motion is also a compound of separate changes in forearm and shoulder angles, as well as changes in other parts of the body.

Consequently, we moved to a coupled differential equation, while retaining linearity. This means that the change in each coordinate and its derivatives is considered to involve corresponding changes in each other coordinate. Here is the equation that we used for each coordinate k:

$$\sum_{j}^{3}[\beta_{jk1}(t)V_{ij}(t) + \beta_{jk2}(t)A_{ij}(t)] + J_k(t) = f_{ik}(t) \tag{5}$$

The unknown parameters in this equation are the 6 weight functions per coordinate k, namely the $\beta_{jk1}(t)$'s and $\beta_{jk2}(t)$'s. We extended the principal differential analysis method of Ramsay (1996) to this case in order to estimate these functions. These are estimated by minimizing the error criterion

$$Q_k(\beta) = \sum_{i}^{123} \int f_{ik}^2(t)\, t. \tag{6}$$

Figure 8, 9 and 10 plot the mean forcing function $\bar{f}_k(t)$ for each coordinate, and, for comparison purposes, the mean jerk $\bar{J}_k$. The 95% pointwise confidence bands on $\bar{f}_k(t)$ indicate that the forcing functions tend to be near 0 in size, so that a homogeneous version of equation (5) might well model the data satisfactorily.

The fit of the equation to the data can be explored by solving the homogeneous differential equation $L(t)u(t) = 0$ for each coordinate, where the linear differential operator L is the left side of (5). The family of solutions to this equation is six dimensional in the sense that there exist six linearly independent functions $u_{jk}(t)$ for each coordinate k. Any solution is a linear combination of these *null space* functions. We may then approximate each of the 123 actual cycle curves or their derivatives by regressing these on these functions. In the special case of fitting position $P_{ik}(t)$, we must also add a constant function.

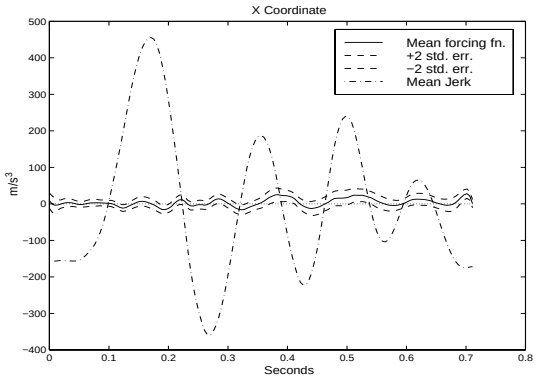

Figure 8: The solid line close to zero is themean forcing function for the X coordinate. The dashed lines on either side are 95% pointwise confidence limits, and the dashed-dotted line is the mean third derivative, displayed to indicate the relative size of the forcing function.

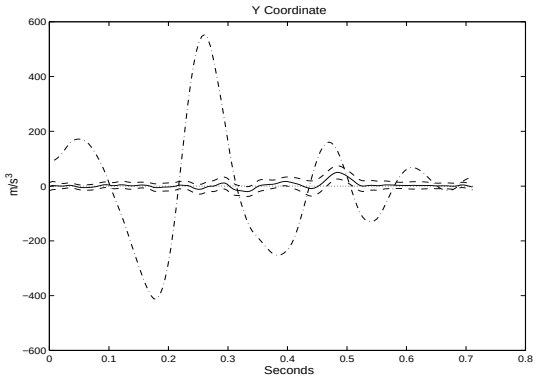

Figure 9: The forcing function for the Y coordinate.

Figure 11 shows how well the velocity curves $V_{ik}(t)$ are fit for the first cycle, which shows a fairly typical level of fit. This quality of approximation is also about what is also achieved for both position and acceleration. Thus, the equation does a fine job of capturing both the average curve shape, and also its variation from cycle to cycle.

7 Discussion and Conclusions

The differential equation seems to work well. This confirms that the body is behaving essential like a closed mechanical system, described in physics by a second order linear equation

$$\beta_0 P(t) + \beta_1 V(t) + A(t) = 0 \tag{7}$$

but with the additional feature that the spring constant, β_0, determining the period of oscillation, and the damping coefficient β_1 are constantly changing

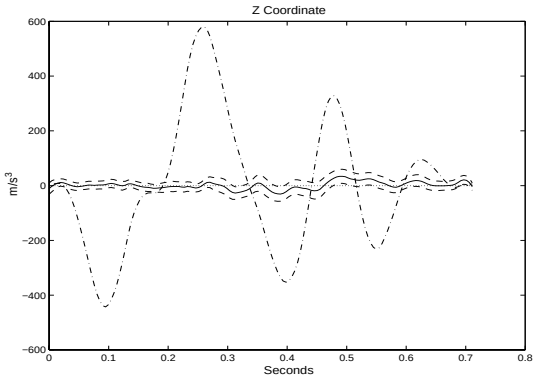

Figure 10: The forcing function for the Z coordinate.

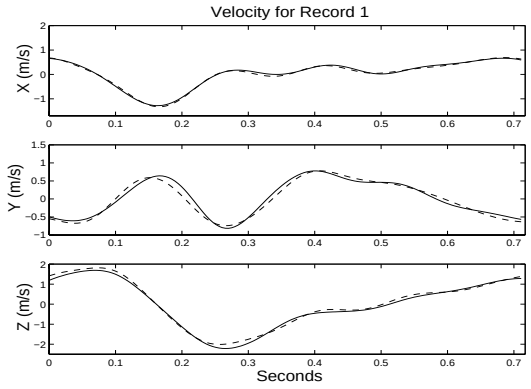

Figure 11: The actual velocity coordinate functions $V_{ik}(t)$ are plotted as dashed lines, and the fit based on the homogeneous linear differential equation of order two is plotted as a solid line. The data are for the first record.

thorugh time. These changes are a result of the neural activation of the many muscles controlling this complex system.

Moroever, we were struck by the strongly periodic behavior of acceleration, visible in Figure 5. The period is about 120 msec, and corresponds to what we observed in the handwriting study. This highly bursty nature of acceleration seems to reflect the character of neural activation, and evokes the concept of a clock time characterizing activity within the motor control area of the brain. It also corresponds to rates of event productiion achieved by experts in other areas, such as speech, typing and playing musical instruments.

It may be that we can achieve better results, or similar results more parsimoniously, by searching for a more natural coordinate system. We are currently exploring an elliptical coordinate system on the horizontal plane, an approach that requires the estimation of a natural point of origin.

There are many statistical issues that remain to be addressed. Can we find a way to simultaneously estimate derivatives and the weight coefficient functions, rather than splitting the process into two as we have done? Can we incorporate some stochastic features into the model, such as random coefficient functions? Can we develop some inference tools to indicate when we should allow for features such as nonzero forcing functions f_{ik}?

The statistical problems associated with estimating differential equations from data seem not to have attracted much attention, except for classical time series contexts. We hope that our results will attract new research in this exciting area.

References

Ramsay, J.O. (1996) Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society, Series B,* **58,** 495-508.

Ramsay, J. O. (1998) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B*, **60**, 365-375.

Ramsay, J. O. (1999) Functional components of variation in handwriting. *Journal of the American Statistical Association,* to appear.

Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data Analysis.* New York: Springer-Verlag.

Ramsay, J.O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B.,* **60,** 351-363.