

Introduction to Functional Data Analysis

DANIEL J. LEVITIN
McGill University

REGINA L. NUZZO
Gallaudet University, Washington, D.C.

BRADLEY W. VINES
University of California at Davis

J. O. RAMSAY
McGill University

Abstract

Psychologists and behavioural scientists are increasingly collecting data that are drawn from continuous underlying processes. We describe a set of quantitative methods, Functional Data Analysis (FDA), which can answer a number of questions that traditional statistical approaches cannot. These methods are applicable for analyzing many datasets that are common in experimental psychology, including time series data, repeated measures, and data distributed over time or space as in neuroimaging experiments. The primary advantage of FDA is that it allows the researcher to ask questions about *when* in a time series differences may exist between two or more sets of observations. We discuss functional correlations, principal components, the derivatives of functional curves, and analysis of variances models.

Introduction to Functional Data Analysis

New types of data require new tools for analysis. Our aim in this paper is to introduce functional data and the methods that have been recently developed to analyze these data. Although functional data themselves are not new, a new conception of them is necessary as a result of the increasing sophistication of data collection in the behavioural sciences. Data collection technology has evolved over the last two decades so as to permit observations densely sampled over time, space, and other continua; these data usually reflect the influence of certain smooth functions that are assumed to underlie and to generate the observations.

Classical multivariate statistical methods may be applied to such data, but they cannot take advantage of the additional information implied by the smoothness of the underlying functions. The functional data analysis (FDA) methods that we describe here can often extract additional information contained in the functions and their derivatives, not normally available through traditional methods.

The goals of this paper are to survey some of this methodology rather than to offer a detailed account, to provide a conceptual and intuitive introduction, and to provide links to further resources (e.g., Ramsay & Silverman, 2002, 2005; Vines, Nuzzo, & Levitin, 2005). We aim to show in broad strokes the kinds of data that can be treated, and the kinds of questions that can be answered within an FDA framework. We will illustrate how FDA works in an interesting experimental context on music and emotion that brings together much of the novelty and power of FDA (Vines, Nuzzo, Krumhansl, Wanderley, & Levitin, 2006). We chose this example because very different data from studies in dozens of other domains can be analyzed under a quite similar approach to that taken in our illustration. We do not assume any particular level of background for readers of the present article. We have attempted to make this introduction intuitive, rather than mathematical (with references to more mathematical treatments for those who are interested), and we have attempted to make this article conceptual rather than a "step-by-step" guide about how to perform FDA (with references to such resources also included). We introduce a few equations to establish a mathematical basis for some of the concepts, but these require nothing more complicated than first year calculus, and nonmathematical readers can skip these with no loss of comprehension of the larger concepts we are trying to describe. By the end of this article, we hope that readers will be able to know whether or not some of their own data may be considered *functional data*, what advantages they may see by analyzing their data using FDA, and finally, where to look next to begin conducting such analyses.

What are Functional Data?

A functional datum is not a single observation but rather a *set of measurements* along a continuum that, taken together, are to be regarded as a single entity, curve or image. Usually the continuum is time, and in this case the data are commonly called "longitudinal."

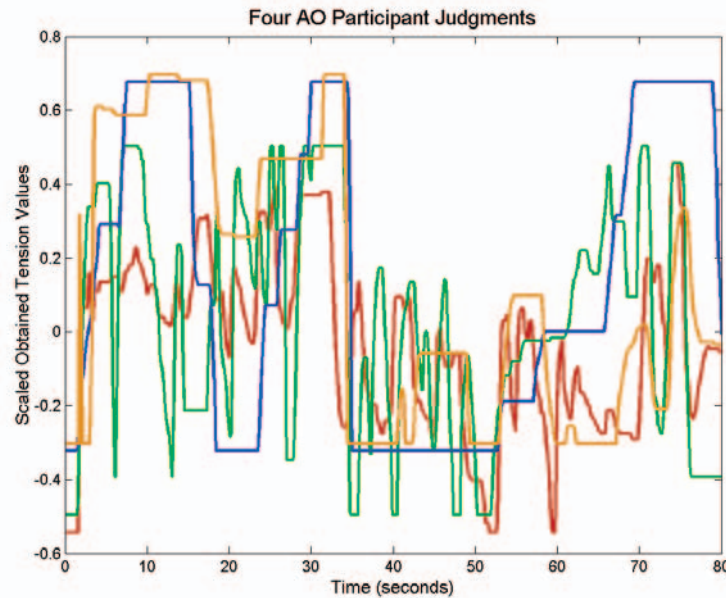


Figure 1. An example of functional data: Tension judgments made by four typical participants in the “auditory only” condition of the experiment described in the text. Each line represents one participant’s judgment over the 80-second duration of the musical performance.

But any continuous domain is possible, and the domain may be multidimensional. In neuroimaging, for example, the activation levels observed at voxels in the brain are the responses over the two or three dimensions of space and possibly time as well. Or, data may be distributed over a continuous psychophysical space (such as vowel space, musical pitch space, or colour space).¹ In this paper, we deal primarily with

1 By “psychophysical space,” we mean that model of human perception or mental representation that is derived from psychophysical experimentation, and is often represented as a manifold in n dimensions, and which does not necessarily possess the same dimensionality as the corresponding physical stimuli that gave rise to it.

For example, the *physical* representation of colour hue can be thought of as one-dimensional: As wavelength decreases, our perception of colour passes through the familiar rainbow from red to orange to yellow, etc. But our psychological representation is actually two-dimensional – circular – as evidenced by experimental findings that colours at the opposite ends of the physical continuum, red and violet, are judged to be more similar than colours that are closer together on the physical continuum such as red and green. Thus, the psychological (or psychophysical) space is two-dimensional and continuous, and data collected from judgments within this space could be considered functional in (psychophysical) space. Similarly, judgments of vowel similarity (Shepard, 1972; Sinnott, Brown, Malik, & Kressley, 1997), musical pitch (Krumhansl, 1990; Shepard, 1982), and musical timbre similarity (McAdams, Winsberg, Donnadieu, & De Soeta, 1995; Plomp, 1970) suggest three-dimensional representations.

functional data measured over time, but the techniques described can be applied to other domains.

Many active domains of current research in psychology and the behavioural sciences produce functional data. Psychobiological functions that change over time such as the levels of hormones in the blood, emotions, or the number and intensity of depressive episodes, are just a few of many examples. Human communication – including spoken and signed language, music performance and perception – is another domain in which the data evolve over time, and the researcher may wish to test hypotheses about the time at which certain events, or responses to those events, occur. Boker, Xu, Rotondo, and King (2002) note that human communication violates the assumption of *stationarity* (i.e., for which consecutive values are dependent over time in a nonpredictable way; for further discussion of stationarity see Box, Jenkins, & Reinsel, 1994; Hendry & Juselius, 2000; Shao & Chen, 1987) and suggest that understanding this nonstationarity may be crucial to understanding the phenomenon. We believe such cases can be profitably understood through the application of FDA, and that the nature of the resulting functions can provide crucial information for understanding human behaviour.

Figure 1 displays many of the features of functional data, and the experiment giving rise to them will be used throughout the paper as a source of illustrations. The aim of the experiment was to quantify the relative

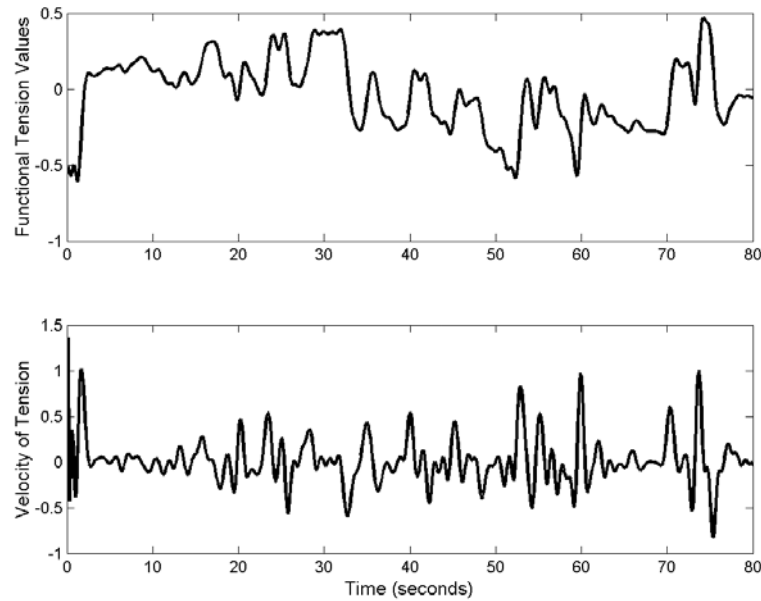


Figure 2. Top panel: Tension judgments for a single participant from Figure 1. Bottom panel: The first derivative of the tension judgments shown in the top panel.

roles of auditory and visual information in conveying emotional and structural aspects of music. The experimental participants were presented with a video-taped musical performance under one of three conditions: audio-visual combined (the normal way in which a video of the performances would be viewed), audio only, and visual only. The dependent variable was the continuous records of the position of a slider (linear potentiometer) that participants controlled to indicate the *tension* that they were experiencing in real time. It is apparent from Figure 1 that there exists considerable detail in these records, and that they exhibit some features that are stable across participants, as well as variation that is participant-specific (including interparticipant differences in motor action planning, interpretation of the task, and decision variability), and undoubtedly some high frequency variation that we will treat as experimental error resulting from nuisance variables such as measurement error, random factors, etc.

One traditional approach to data analysis would require that we obtain an average for each of the three conditions over the time course of the experiment, and compare those means. Yet it is apparent from Figure 1 that such an average would obscure interesting time-based variations in the ratings. Using an ANOVA, we could answer the simple question about which of the three conditions yielded the highest ratings, and with either planned orthogonal contrasts or

post-hoc tests, we could ascertain which means, if any, differed from the others. A second traditional approach might be to treat each observation as a repeated measure, and conduct an RM ANOVA or MANOVA, neither of which is satisfactory because of the obvious autocorrelation between successive points: With the linear slider used by participants, observations are not independent because the slider must pass through intermediate points on its way from one desired position to another; the data violate stationarity. In addition, there are other factors to consider in such data, such as regression to the mean, and the potential underlying phasic attributes of the musical piece being judged. Finally, traditional GLM-based methods would not allow for the precise comparison of interparticipant responses, which may also be of interest. Although one could conduct correlational analyses among participants, factors such as differential reaction times and differential use of the scales could obscure relations in the data. Thus, traditional statistics would allow us to answer only simple questions, and then, only with certain independence assumptions violated.

To begin the FDA approach, we make the assumption that the underlying process generating the data is smooth. The tension in music experienced by a listener rises and falls in response to features in the music that evolve over time, and typically without abrupt, instantaneous changes. The observed data are subject

to measurement error and other types of local disturbances that may mask this smoothness, but FDA includes techniques to identify the smooth sources of variation by filtering out the unwanted variability. Other examples of processes assumed to be smooth include the depression level of a patient over time, the evoked response potential (ERP) measured on the scalp with surface electrodes, finger tapping in response to mental imagery or synchronously with an auditory stimulus, body core temperature, and the flow of blood through a region in the cortex in response to event-related tasks.

The continuous underlying process may be observed not only at multiple time points (the discrete observations which, in our music data, are taken 10 times per second), but at multiple manifestations as well (the tension curves are replicated across participants). FDA was designed to take advantage of replications, and in particular those produced by controlled experiments.

When we say that a curve is *smooth*, we usually mean that it is differentiable to a certain degree, implying that a number of derivatives can be derived or estimated from the data. These derivatives, especially the first two, “velocity” and “acceleration,” often have interpretations relevant to the study, and thus analyzing these derivatives can be an important component of FDA (Vines et al., 2005). For example, we might want to know the rate of change in tension in response to a particular musical event, in addition to the level of tension. Useful derivative estimates require both a number of sampling points and a signal-to-noise ratio that is sufficiently high, features that are not often found in the kind of longitudinal data that are subjected to hierarchical or multilevel linear modeling.

Derivatives are used in other ways in FDA, and the exploitation of derivative information is a central theme. Figure 2 shows the functional observation of tension for a single participant; on the top, the discrete raw data are shown as points and an estimate of the smooth underlying tension curve is indicated by a continuous line. In the lower panel, the first derivative of this curve estimate is displayed. The use and interpretation of derivatives in functional data is a rather new, specialized topic in FDA (see Ramsay & Silverman, 2005; Vines et al., 2005). Intuitively, one can imagine the ways in which such derivative information can reveal patterns in a (functional) dataset that address important research questions. In the case of depression levels in patients, one might want to know not just when they reach a peak, but how rapidly they rise and fall (the first derivative). In studies of regional cerebral blood flow (rCBF), one might want to know not just which neural structures are being

supplied by blood at a given point in time, and not just how fast the blood is traveling from one region to another, but whether the blood flow *rate* shows significant acceleration or deceleration in response to particular tasks.

An Overview of Functional Data Analyses

FDA extends the capabilities of traditional statistical techniques in a number of ways. Studying the variability of a dataset when the observations are curves instead of points requires that standard tools be adapted for this functional framework, and that new tools be created to take advantage of the unique characteristics of functional data. For example, the availability of derivatives permits the use of differential equations as models, and thus introduces descriptions of the dynamics of processes as well as their static features. Since the underlying processes one studies with FDA are, by definition, changing over time (or space, or some other continuum), the ability to quantify these dynamics is a principal advantage of these techniques over traditional static techniques.

A comparison of FDA with traditional multivariate statistics (MVA) is valid at many levels. In classical statistics, a sample of discrete points is taken from a large population. This sample may contain many sources of variation: Initial variability across the units being measured, measurement error, variability due to sampling, inability to reproduce the treatment conditions exactly from one unit to another, and the systematic variability that is the goal of the analysis. In a functional setting, a sample of *curves* is taken from a larger population of curves. This functional sample contains the same sources of variation as does a discrete sample; now, however, there is the additional challenge of quantifying variability within and across curves as opposed to discrete points.

The relationship between FDA and models for repeated measures or longitudinal data of the mixed effects and structural equation varieties is even closer. As a rule, FDA is concerned with data able to support the estimation of more complex curves than is usually the case with these older approaches. FDA data tend to have a higher signal-to-noise ratio in each observed value, and to be observed over a number of sampling points ranging from 20 or 30 to tens of thousands or more. As a result of this greater resolving power, the estimation and use of information in derivatives tends to be central to FDA. (For reviews of multivariate methods for longitudinal data see Cudeck, 1996, and Pinheiro & Bates, 2000; Ramsay, 2002 provides further detail on the relationship between FDA and other longitudinal data approaches.)

A typical analysis of functional data begins by using

nonparametric smoothing techniques to represent each observation as a functional object. The original data are then set aside, and the estimated curves themselves are used for the rest of the analysis. Further processing of the curves is also possible, such as the taking of derivatives or performing transformations.

These functional objects are next usually represented in a Cartesian coordinate system with time along the x-axis and the value of the dependent variable along the y-axis. At this point it may be apparent that there are two sources of variability in the curves that comprise the functional data, each one corresponding to one of the dimensions of the graph of the curve: “Amplitude variability” having to do with variation in the shapes of the curves as reflected in peaks and valleys (along the y-axis), and “phase variability” having to do with shifts in the timings of these features (along the x-axis). If appropriate, this amplitude and phase variability can be extracted and removed through scaling and curve registration techniques, respectively, after which these sources of variation can themselves be further analyzed and related to one another, or to create individual difference profiles for experimental units or participants.

As with traditional techniques, the *exploration* of functional data is often preliminary to more structured analyses, and we have found that plotting observed effects is a crucial stage in the interpretation of these analyses that can greatly aid in choosing further analyses (cf. Cleveland, 1993; Tufte, 1983; Tukey, 1977). Given a single functional variable, we may want to estimate the correlation between curve values taken at two time points s and t ; the result is a smooth bivariate correlation function $r(s, t)$ that is the analogue of the correlation matrix in MVA. Principal components analysis (PCA) is as valuable in FDA as it is in MVA for exploring the structure of the correlation function, and PCA has been adapted to permit the smoothing of estimated functional principal components.

A *functional linear model* may involve functional data on either side of the equation, and usually in conjunction with multivariate observations. For example, a functional analysis of variance (fANOVA) involves a functional response modulated by a set of experimental conditions, and in this context it can be useful to view indicators of effects such as F -ratios and squared multiple correlations as themselves functions of time or whatever other continua are involved. This introduces interesting new versions of the multiple comparison problem, and solutions are presently under development.

Linear models involving functional covariates require the estimate of regression *functions* as opposed to MVA's regression coefficients. Functional indepen-

dent variables also introduce the need for imposing smoothness on their regression functions because a functional covariate can represent a number of degrees of freedom for fit that exceeds the sample size, and this could in principle produce over-fitting of the response data.

When a data set involves two or more variables, investigators are usually interested in how they co-vary. The cross-correlation function $r(s, t)$ then consists of estimated correlations between one set of function values at time s and the set of functions values at time t . This information can be displayed graphically, but we usually want to supplement such displays by methods that look for more specific correlational structures. Canonical correlation analysis is perhaps more useful in FDA than it tends to be in MVA, in part because canonical variables can be smoothed to avoid having them reflect too much uninteresting high-frequency variation.

Boker et al. (2002) describe an interesting example of functional data in which windowed cross-correlations are used to identify linkages between functional variables that may be specific to certain time intervals that may be lagged by an amount that must be estimated from the data. Their experiments involving gestures and body movements for couples during conversation and dance are compelling illustrations of the spirit of FDA at work. We now consider the stages of FDA in more detail.

Applying FDA

Discrete Functional Data

We use the notation $x(t_j)$ for the value of a continuous underlying process at time t_j , and the notation y_j to indicate a corresponding noisy observation of the process $x(t)$, where $j = 1, \dots, n$. A single functional observation therefore consists of n pairs (t_j, y_j) , and we want to use these data to estimate characteristics of $x(t)$ such as function or derivative values at unobserved time points.

Replications of this underlying process may be available. Again, by replications we mean either multiple samples of a single process (improving our estimates of the actual shape of the curve that represents the underlying process), or multiple participants providing data on the same dependent variable(s). Earlier, we pointed out that FDA is intended to be used on data that come from a smooth and continuous process. However, virtually all data collection that we know of produces noncontinuous observations, taken as samples at discrete points in time. Even when the sampling rate is very high (as in audio recordings routinely made at 44 KHz or even 192 KHz sampling rate), the points, technically, are not continuous. Thus, when we

refer to discrete observations, the assumption is that enough observations exist to model the underlying process (cf. Nyquist, 1928; Shannon, 1949), and this is largely a matter of experience, experimenter intuition, and trial-and-error. Thus, because FDA attempts to model a smooth process from discrete points, it is generally inappropriate for small datasets, and its greatest advantages over traditional methods of analysis occur with especially large datasets.

The i th replication of the underlying process is written as $x_i(t)$, and data values corresponding to time points t_{ij} are written as y_{ij} (and where t can take on any real values). Often the observations are taken at the same arguments t_j for each replication; in this paper, we will assume for notational simplicity this is true, that is, that $t_{ij} = t_j$ for each i , although in actual applications the locations and even the number of time points can vary from replication to replication. This might occur, for example, in an experiment in which a dependent variable was to be observed at precise points in time, but one or more measurements were taken somewhat off-schedule. This happens *often* in clinical settings. Imagine one wanted to measure blood serum concentrations of a particular pharmacological agent and blood samples are scheduled to be drawn at 60-minute intervals. A given sample for a particular participant or trial might be drawn some minutes earlier or later than the nominal time due to factors external to the experiment or due to experimenter error (e.g., Nuzzo, 2002). Such variability can be handled with techniques described in the section “Registration of Functional Objects” below (see also Ramsay & Silverman, 2005).

Although measurement error usually does exist in the observations y_j , we will assume for the moment that the t_j are measured without error. We can relate the discrete observations to the underlying smooth process through the equation $y_{ij} = x_i(t_j) + \epsilon_{ij}$, where each ϵ_{ij} is the disturbance or error term, assumed to be distributed with mean zero and finite variance σ^2 .

Building Smooth Curves From Discrete Data

The first step in a functional data analysis is to convert the raw data into functional objects. To do this, a curve is fit to the discrete observations, which approximate the continuous underlying process. Then the discrete points are set aside and the functional objects retained for subsequent analyses.

A common procedure in statistics, mathematics, and engineering for representing discrete data as a smooth function is the use of a *basis expansion*.

$$x_i(t) = c_{i1} \phi_1(t) + \dots + c_{ik} \phi_k(t) + \dots + c_{iK} \phi_K(t), \quad (1)$$

where K indicates the number of basis functions and $f_k(t)$ is the value of the k th basis function at argument value t . The basis functions $f_k(t)$ are a system of functions specially chosen to use as building blocks to represent a smooth curve. There are many different types of basis function systems, of which the best known are powers of t or monomials, used to construct polynomials, and the Fourier series used for periodic curves. The c_{ik} are the *basis coefficients* and determine the relative weights of each basis function in constructing the built curve for curve i . Estimating a curve that is expressed as a basis function expansion then reduces to the essentially multivariate parameter estimation problem of estimating these coefficients.

When curves are not periodic and are complex, so that a low-order polynomial cannot capture their features, the most commonly used basis functions are the B-spline functions (de Boor, 2001; Shikin & Plis, 1995; Unser, 1999), and these will be used in this paper to represent the tension curves in our illustrative music and emotion study. B-splines are essentially polynomials joined end-to-end at a set of interval boundaries called *knots*. This structure allows for differing amounts of smoothness or roughness at various places along the curve, that is, each strand between knots can be parameterized in an optimal fashion. Knots are often chosen to be equally spaced, but they may also occur at the exact times the data were actually observed, or at prespecified time points of interest. B-splines are characterized by their *order*, which is one larger than the degree of the polynomials from which they are constructed; for instance, B-splines of Order 3 comprise quadratic curves joined at the knots, and B-splines of Order 2 are constructed from piecewise lines (i.e., line segments). A set of B-splines, composed of a B-spline which has been replicated and shifted down the number line, can serve as basis functions. Across a wide range of datasets, it has been found that Order 4 B-splines appear reasonably smooth to the eye and thus they are often used in standard smoothing packages.

B-splines are normalized so that for any fixed value t , the sum of all B-spline basis functions at that point is one. This normalization means that a coefficient c_{ik} is approximately the value of the curve being fit at the location where the k th B-spline basis function achieves its maximum.

The basis coefficient weights are chosen so that the constructed curve will optimally fit the data for a certain degree of smoothing. The amount of smoothing is controlled by the number of basis functions used: The greater the number of basis functions K is, the more the built curve will exactly fit or interpolate the discrete points; the smaller K is, the more the curve is a

smoothed version of the points, but at a cost of a reduced capacity to capture sharper features in the curves.

There is a decision associated with choosing the number of basis functions, often framed in terms of the *bias-variance trade-off* common in many statistical analyses. A high number of basis functions will yield a curve that is more faithful to the observed data (low bias) but that is often less smooth (high variance). Using a small number of basis functions will produce a curve that places less importance on interpolating the discrete points (high bias) but more importance on smoothness (low variance). Put another way, under-smoothing of the curves leaves in artifacts and variability (“noise”) that are not truly part of the process being observed; the resulting curve may thus represent perturbations in the observation process that do not pertain to the analysis being sought or the research question being asked. On the other hand, over-smoothing discards small-scale, high-frequency behavioural data that may be part of the process we wish to observe and analyze – maybe even the most interesting part! Thus, there is an art associated with building a curve through basis expansion, and the analyst should cautiously experiment with various parameters, driven by theories about the underlying process whenever possible. Finally, a data-driven approach to basis function specification is also possible; functional principal components analysis, described below, can be viewed as a method for constructing an optimal orthogonal basis of fixed dimensionality, and these are often referred to as *empirical basis functions*. Automatic methods for basis expansion is an area of active research (e.g., Koenker & Mizera, 2004), but in their present state of development should not take precedence over informed and careful judgment.

Another way to build a smooth curve from discrete data is through the use of a *roughness penalty* (PEN below). Here we have direct control over the amount of smoothing, and the problem is formulated as a direct trade-off between fitting the data and producing a smooth curve. We often fit the data using as our goal the minimization of the least squares criterion $SSE = \sum_j \{y_j - x(t_j)\}^2$. The smoothness of the fitted curve can be quantified using the criterion $PEN_2 = \int D^2 x(s)^2 ds$, which many readers will recognize as the integrated squared second derivative. This penalty describes the total curvature in the curve x , and is popular in the smoothing literature (Simonoff, 1996). The smaller the size PEN_2 is, the closer to a straight line the curve will be. Of course, these two criteria, data fit and smoothness, are normally in conflict, and we can put them together to create a penalized residual sum of squares

$$PENSSE_\lambda = SSE + \lambda \text{ PEN}_2. \quad (2)$$

The smoothing parameter λ controls the amount of importance we place on each of the two competing goals: As λ approaches infinity, the curve approaches the straight line obtained from a linear regression; as λ approaches zero, the curve approaches a direct interpolant of the data, with cubic polynomials connecting the observed points. The roughness penalty method makes explicit the bias-variance trade-off explicit, and allows us to directly quantify our priorities through the choice of λ . Again, automatic methods exist for choosing the best value of λ , the most popular of which involves a technique of partitioning the data called cross-validation (Ramsay & Silverman, 2005). However, knowledge of the process generating the data and a spirit of experimentation often prove to be the most important tools in the area of spline smoothing (as they often are in all data analysis).

The use of roughness penalties is not limited to the creation of a functional object from discrete data. The idea of making explicit the competing goals in an analysis is a powerful one in the field of statistics. Green and Silverman’s (1994) book contains many examples of applications of this idea. Since they lend themselves naturally to working with curves and smoothness, roughness penalties are also very helpful for adapting the tools of classical multivariate statistics to an FDA framework.

Basis function expansions, whether roughness penalized or not, are by no means the only options for smooth curve-fitting, and Fan and Gijbels (1996) can be consulted for information on *local polynomial smoothing*, for example.

Registration of Functional Objects

Consider a case in which we graph our dependent variable (on the y-axis) as a function of time (on the x-axis). As mentioned above, it is common in functional data analysis to speak of two types of variability in the observed curves: amplitude variation, which deals with the differences in height between the curves, and phase variation, which deals with the differences in timing of important features between the curves. For example, pure amplitude variation might occur when two individuals react simultaneously to the same musical event, but differ in the intensity of their reaction, giving rise to two curves of different heights (but with temporally synchronized peaks). Pure phase variation might occur when the individuals react with the same intensity to a specific musical event, but differ in *when* they react, giving rise to curves that are temporally misaligned (but with concordant amplitudes).

The possibility of phase variation is characteristic of

functional data. Often in functional data, it is the amplitude variation that is of primary interest, since this describes how the heights or intensities of the curves vary among participants (or among any groups that may exist), and in this case it can be critical that phase variation is first removed. In other circumstances, the phase variation is actually the measure of interest and can be analyzed itself. For example, one researcher may be interested only in the extent of responses, and regard variation in the timings of reactions to be a nuisance (the result of differing decision times, reaction times, and neural clock speeds among participants); one would then seek a technique to correct event times for certain participants so that all reactions are simultaneous. Such an approach would be consistent with the notion that the dependent variable is tracking an underlying process that has peaks and valleys at certain objective time points, with the timing variations representing simply individual differences in reaction time that are not of relevance to the study. Another researcher, however, may wonder whether the phase variation is itself due in part to the experimental manipulations, and want to study this issue. In either case, the separation of these two curve characteristics in some way is required.

It is helpful to think of time as being a variable that can be manipulated. In the case of the tension ratings, clock time can be different than reaction time, and we would usually like to transform the clock time for each rater to reflect a standardized biological time. In time-warping registration, the idea is to scale, by locally shrinking or expanding clock time, the moment each response was recorded for each subject so that it conforms with a standardized biological time. In some simple situations, this may be a matter of a simple linear transformation of the timings for certain individuals. But often a more complex *time-warping function* is required that stretches time over some intervals and compresses it over others.

We call our time warping function $h(t)$. When a process is running faster than the standardized time, $h(t) > t$; if it is running more slowly, $h(t) < t$. If our original curve is $x(t)$, the registered curve would be $x[h(t)]$. Mathematical procedures have been developed to estimate time-warping functions for a dataset, and they have the following characteristics. First, we normally expect the time-warping function to leave the beginnings and ends of the data curves intact. That is, $h(t_0) = t_0$ and $h(t_{\text{end}}) = t_{\text{end}}$. Secondly, we impose an assumption of monotonicity (an ordering assumption) in that we require that the events in the registered curves occur in the same order as those in the unregistered curves. In mathematical terms, we would require then that $h(t_2) > h(t_1)$ IFF $t_2 > t_1$, which

implies that the time-warping function is monotonically increasing. Lastly, we have the goal that our registered curves all possess the same shape as the original curves – that is, only the amplitudes are different, but the peaks and valleys occur at the same time. This would mean that the registered curves in the dataset are proportional to one another. In mathematical terms, we say that if curves x_1 and x_2 are proportional then $x_1(h(t)) = a x_2(t)$ for a positive constant a .

To estimate the time-warping functions for a dataset, we again call upon the idea of roughness penalties. As before with the curve estimation, we have two competing goals in registration: On the one hand, we want to line up all the features of the curves, and on the other hand, we do not want the time-warping function to be too rough, because taken to an extreme this would be “over-registering” the curves in the same way that it is possible to “over-fit” a model to data. In order to balance these two goals in our registration procedure, we need quantitative measures of each. For the first registration goal, researchers have developed a “misfit” measure MF, quantifying the amount of misalignment between features of the curves, which is analogous to the sums of squared errors in curve estimation (Kneip, Li, MacGibbon, & Ramsay, 2000; Ramsay & Li, 1998). To quantify the second goal of a smooth time-warping function, a standard roughness penalty R is used, which is usually the integrated squared second derivative of the time-warping function h . Again as in curve estimation, the estimation of the time-warping function is achieved by minimizing the overall criterion $MF + \lambda R$, where λ is the smoothing parameter.

The registration procedure requires a standardized curve to which the observed curves will be aligned. In practice, a “gold standard” curve is often not available, and so one must be constructed from the data. A good method is an iterative one: Find the mean function of the original curves in the dataset; register the original curves to this mean curve; find a new mean function of these registered curves; and repeat. Usually only one or two iterations are necessary, and registering the curves to the initial mean function may even be satisfactory. As in traditional data analysis, a good rule of thumb is to inspect the results at each stage to make sure they make sense! The experimenter may have some knowledge about the underlying process that the curves are intended to represent, and this can inform smoothing and registration operations. For example, if the experimenter knows that two peaks of activity existed in the original stimulus set at particular points in time, she or he would obviously want to avoid any registration solution that collapses those two peaks into a single peak.

Adaptation of Classical Methods for Functional Data

Functional analogues for many multivariate statistical methods have already been developed and published in the FDA literature (Guo, 2002, 2003; Vidakovic, 2001). These include simple linear models, ANOVA, generalized linear models, generalized additive models, clustering, classification, principal components analysis, and canonical correlation analysis. In this section we will describe the functional principal components analysis and the functional linear model; these are good representative examples of adapting multivariate methods to the functional framework.

Functional Principal Components Analysis

Principal components analysis (PCA) is a data reduction technique conducted when one wishes to a) reduce the number of (possibly) correlated variables to a smaller number of uncorrelated variables, and b) reveal latent structure in the relations between variables. PCA finds the linear combinations of variables that highlight the principal modes of variation.

For functional data, on the other hand, the principal components are defined by PC weight functions varying over the same range of t as the functional data. The PC weight functions serve the same purpose as the multivariate principal components: to characterize the principal types of variation among the observed curves. These components can also be rotated to enhance interpretability. Each unrotated component is associated to an eigenvalue that indicates the variance in the modality that the component defines, and the sum of these eigenvalues indicates the total variability accounted for by the associated PC weight functions. Principal component scores for each curve can also be computed for each rotated or unrotated principal component. Thus, the only critical change in moving from multivariate to functional PCs is that the components are functions rather than vectors.

Functional principal components analysis may also incorporate a roughness penalty into its methods in order to ensure that the estimated principal components are sufficiently smooth. A common roughness penalty is the integrated squared second derivative of the principal component curve; this penalty was also used in the section on curve-fitting.

As with principal component vectors in classical PCA, we may wish to rotate the principal component functions to increase interpretability. A functional analogue has been developed for the classical VARIMAX rotation strategy, which finds principal component curves whose values have greater variability about zero. However, as in multivariate analysis, there is no

essential requirement that the rotation maintain orthogonality. Other tools of classical PCA can also be used in the functional case; scree plots are helpful to determine the number of principal component curves to retain, and plotting scores of the participants on the first two principal components can highlight hidden types of subject groups.

Finally, we may choose to apply functional PCA to various derivatives or mixtures of them in order to study variation in, for example, velocities and accelerations.

Functional Linear Models

Most aspects of the multivariate linear model are readily applicable to the case of functional data. This section will discuss simple linear models, but the same tools can then be generalized to create functional nonlinear models, functional generalized linear models, and so on. In the functional setting, either the response, the predictors, or both variables are curves.

The simplest case is when the response $y_i(t)$ is functional, and predictors x_{ij} are multivariate. In this case the model is

$$y_i(t) = \sum x_{ij} \beta_j(t) + \epsilon_i(t), \quad (3)$$

and we note that this is simply a conventional regression analysis for each value of t . Consequently, the regression coefficients $\beta_j(t)$ are functions of time. These regression coefficient curves show how the effect of the predictor changes the response at each time t . In fact, the conventional univariate regression problem simply describes how the predictor affects the response in the snapshot of time when the data were collected. However, a new feature in the functional situation is the possibility of forcing the regression functions to be smooth by adding one or more roughness penalties to the usual least squares criterion, in much the same manner as we did in (2). We will apply this form of the functional linear model to the tension data in Section 3, where we discuss the effects of experimental conditions on the tension curves.

Functional predictors $x_i(t)$ with a scalar response y_i are also possible. Here the coefficients are also curves, and they describe the contribution (at every point in time) of the pattern of independent variables toward the scalar dependent variable. The model looks like this:

$$y_i = \int x_i(s) \beta(s) ds + \epsilon_i \quad (4)$$

where $\beta(s)$ is a regression coefficient function and ϵ_i is a disturbance or error term. The integral of the prod-

uct replaces the sum of products in the classical linear model.

But a new problem now intrudes here: Since, in effect, each time point is analogous to a new independent variable, implying in turn that we have an infinity of independent variables at our disposal, we have the problem that this model can fit any finite set of data exactly. To keep the model from being trivial and therefore impractical, we must impose a roughness penalty on the regression function $\beta(s)$. By ensuring that $\beta(s)$ is sufficiently smooth, we also can discover whether the model works or not.

The case of functional predictors and functional responses is still more complex: The regression coefficients are then two-dimensional functional surfaces $\beta(s, t)$ rather than one-dimensional functional curves. These surfaces detail how each point in time of the predictor curve $x_i(s)$ affects each point in time of the response curve $y_i(t)$. The model can allow for delayed or lagged influence; for a given point in time, the past pattern of the predictor can influence the present point in the response curve. Future behaviour of the predictor variables may also affect the response curve at any given point; we call such predictors *oracles*. In general, the influence of the predictor can be moderated so that any portion of the predictor curve is included in the linear model. The coefficient surface will reflect the windows of influence chosen. Further details can be found in Ramsay and Silverman (2002, 2005).

Goodness-of-fit diagnostic techniques from classical linear models can be adapted for the functional case as well. The difference is the dependence of these procedures on time t in the functional setting. In a functional residual analysis, the residuals are now functions of time, and a graphical plot of the residual functions that show a trend over time is suggestive of a model with a poor fit. A functional principal components analysis of the residuals may reveal the presence of previously unknown patterns in the data that were not captured by the linear model.

Furthermore, goodness-of-fit statistics can be calculated by using functional analogues of the sums of squares. In the functional response/multivariate predictor situation, the sums of squared errors function, for instance, is

$$SSE(t) = \sum_i [y_i(t) - Z_i \beta(t)]^2, \quad (5)$$

where Z_i is a row vector of covariate values. Similarly, the sums of squares for regression is

$$SSY(t) = \sum_i [y_i(t) - \hat{\mu}(t)]^2, \quad (6)$$

where $\hat{\mu}(t)$ is the overall mean function. From this we can calculate the squared multiple correlation function RSQ with values

$$RSQ(t) = [SSY(t) - SSE(t)]/SSY(t). \quad (7)$$

We can also calculate the mean square error function MSE with values $MSE(t) = SSE(t)/df(\text{error})$, where $df(\text{error})$ is the degrees of freedom for error, or the number of observed curves N less the number of coefficient functions β in the model. Similarly, the mean square regression function can be determined, with values $MSR(t) = [SSY(t) - SSE(t)]/df(\text{model})$. Finally, we can compute the F -ratio function $F\text{-ratio}(t) = MSR(t)/MSE(t)$.

Inferential Statistics

Determination of statistical significance for functional models is an open area of research, and Ramsay and Silverman (2005) discuss a variety of issues and summarize current work.

A simple method for expressing statistically significant regions of regression coefficient curves involves a pointwise comparison of the F -ratio function values to the critical values of the F distribution with appropriate degrees of freedom. Regions in which the F -ratio function exceeds the critical value threshold are statistically significant. This pointwise method is not fully in the spirit of functional data analysis, however, and more appropriate methods might be found in the simulation and resampling of the observed functions.

Differential Equation Models

A central theme of FDA is the use of information in derivatives such as velocities and accelerations. This leads naturally to models for data in the form of differential equations. A differential equation expresses a relationship that holds between a function and one or more of its derivatives. For example, the function $x(t) = e^{kt}$ can also be expressed as the equation $Dx(t) = kx(t)$, where the notation " $Dx(t)$ " means "the velocity or rate of change in x at time t " or "the derivative of x at t ."

Differential equations are standard mathematical tools for physical and biological scientists and for engineers, but seldom appear in behavioural science literatures. Nevertheless, they offer some exciting possibilities as modeling tools. Methods for estimating differential equations such as principal differential analysis (Ramsay, 1996) are a part of the FDA arsenal. See Ramsay (2000) for an example of the use of a differential equation to model some exceedingly complex handwriting data.

Autocorrelation

Let $\{a_i\}_{i=0}^{N-1}$ be a periodic sequence, then the autocorrelation of the sequences is the sequence

$$\rho_i = \sum_{j=0}^{N-1} a_j a_{j+i} \quad (8)$$

where the final subscript is taken modulo N . The autocorrelation is simply the correlation of the sequence against a time-shifted version of itself. The autocorrelation $R_f(t)$ is defined as

$$R_f(t) \equiv \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(\tau) f(t + \tau) d\tau \quad (9)$$

Techniques do not presently exist for accurately calculating the correlation between two functions that are themselves serially correlated and nonstationary (Box et al., 1994, but see Huitema & McKean, 1998 for a discussion of the loosening of restrictions in some cases). In many functional datasets, successive observations of the dependent variable are intercorrelated, as is often the case when the investigator is taking measurements of a continuous variable that, when changing from one value to another, must pass through intermediate values. Examples include neurophysiological measures such as heart and respiration rate, hormone levels or serum blood concentration levels of a substance, or the movement of a human-controlled input device (such as a slider). In data from brain imaging experiments, both the spatial extent and the level of the blood oxygenated level dependent signal are intercorrelated: The likelihood that a given voxel will be activated in an experiment is dependent on whether an adjacent voxel is activated (spatial extent) and the amount of activation at a particular time t is dependent on the activation at time $t-1$. Thus, in all these cases, a given observation is dependent on adjacent observations, violating the independence assumptions in classical MVA.

Alternative windowed cross-correlation methods for functional data that violate the conditions of stationarity are currently in development (by our laboratories and others). Key to our approach is the setting of boundary conditions, that is, computing a correlation in two different ways, one of which is known to overestimate and the other of which is known to underestimate the true degree of relation between two or more variables. Boker et al. (2002) introduced a windowed cross-correlation method that can compute correlations under such circumstances, and that does not assume stationarity. (If we let a vector represent the sequence of observations obtained in an experiment, a subset of contiguous observations sampled from that vector is often called a *window*.) With functional data, one might consider the minimum

and maximum correlation values obtained from such moving window analyses as boundary conditions for the true value of the correlation between the variables of interest. Often with functional data, there are a priori reasons to choose a window of a certain size based on a consideration that it likely contains stable or stationary data due to the time course of the underlying process. Haemodynamic lag, for example, suggests that observations within a time period of 1.5 seconds or so can be windowed for such a cross-correlational analysis.

Bootstrapping

As with conventional statistical analyses, bootstrapping (Efron, 1979; Efron & Tibshirani, 1993) can be an effective tool for providing estimates of the sampling distribution of an estimator by resampling with replacement from the original sample. Two articles on bootstrapping offer a good starting point for the interested reader. McKnight, McKean, and Huitema (2000) introduced a method based on Freedman (1984) for bootstrapping datasets with autoregressive properties. Bootstrapping was recently combined with FDA to evaluate the range of forecasts of sulfur dioxide levels near a power plant (Fernández de Castro, Guillas, & González Manteiga, 2005).

Practicalities

Functional data share characteristics with other types of data such as time series or multivariate datasets. Yet in real-world situations, data can rarely be classified exclusively as one type or another. Some datasets may be unequivocally functional. In these cases, for example, derivatives play an important role in the analyses; data are collected at a very fine resolution; the observed function is very smooth; and many replications of the process are available. Other datasets, when some of those characteristics are missing, may straddle the fence between “functional” and “multivariate” (or “time-series”). Still other datasets may be decidedly nonfunctional. So for most datasets, a variety of tools are available for analysis. Functional data analysis tools are best used when there are too many data points for a repeated measures ANOVA, for instance, when traditional methods ignore correlations between datapoints, or – and perhaps most importantly – when the research question focuses on precisely *when* in a time series (or *where* in a multidimensional psychophysical or Cartesian space) a significant result occurs rather than collapsing across all the observations for a mean or summary difference between conditions. FDA methods are also most helpful when derivatives of the observed functions are important and physically meaningful. Note that the

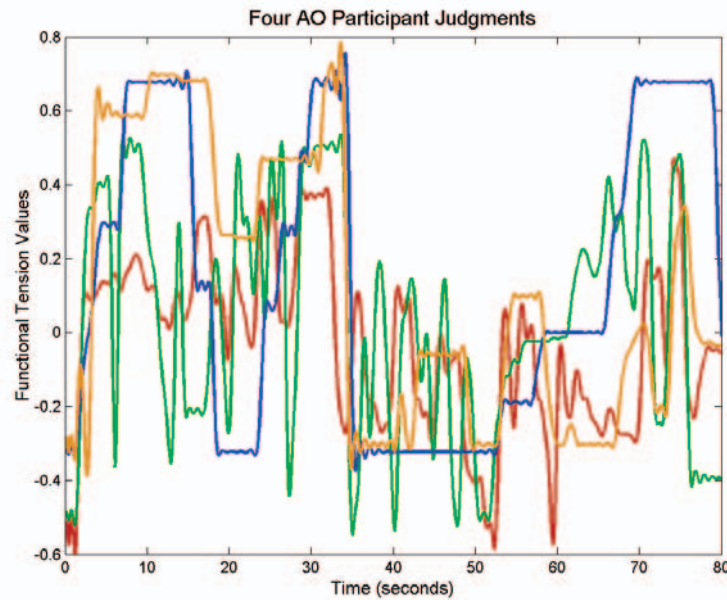


Figure 3. For the same four participants from Figure 1 in the audio only condition, the functional objects obtained after expanding the scaled raw data with 6th order B-splines as described in the text. Note the similarity to Figure 1, indicating the quality of the B-spline estimation.

previously mentioned techniques of Boker and colleagues (2002) can also be used for these purposes.

Software for the analysis of functional data using standard FDA techniques, including those described here, is available for the R, S-PLUS, and Matlab programming languages. The functions with their documentation can be downloaded from the web site www.functionaldata.org. These languages are easily extensible for customized functions and more advanced techniques for researchers with some programming experience. New techniques and software are being developed and shared with other researchers. A background in introductory calculus and statistics is necessary for the basic use of these techniques; familiarity with differential equations, matrix algebra, and more advanced multivariate and nonparametric statistics would be helpful to better understand the concepts. Analysis of functional data using these techniques may require more time of the analyst than it would for multivariate data using standard techniques. Most of this extra time stems from the need to represent the discrete observations with functional curves, using the basis expansion techniques described earlier. This is not a fully automatic procedure; the analyst needs to use prior knowledge and expectations of the process to best represent the data with smooth curves. The rest of the analysis may follow along straightforward lines, if, for example, a functional linear model or functional principal com-

ponents analysis is all that is required. More complicated analyses may require additional programming and time from the analyst.

Case Study: Perceived Tension in Music Over Time

Vines et al. (2006) sought to understand how the physical gestures of professional musicians might contribute to the perception of emotion in a musical performance. Twenty-nine musically trained participants saw, heard, or both saw and heard performances of a Stravinsky concerto by a clarinetist. All participants performed the same task: a continuous judgment of experienced tension in the musical performance, made by moving a slider up and down to indicate minimum or maximum tension. The position of the slider was translated into a discrete value along a continuum from 0 to 127, and was sampled every 10 ms. This example illustrates the ways in which continuous measures (location and time) can be discretized for analysis purposes in FDA. The principal research question of interest was how similar the rated tension experience was for each sensory modality, as a way to understand the similarities and differences in musical tension conveyed by the different sense modalities. A subsidiary research question concerned where in the piece, if at all, the auditory versus visual modalities carried the most influence on the overall experience of tension by an audience member.

The analyses in this section used FDA software

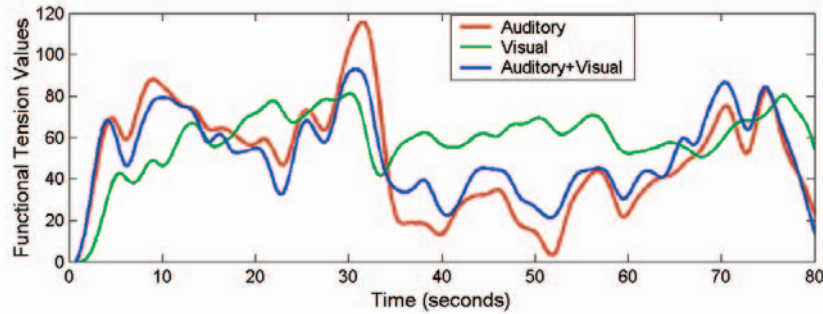


Figure 4. Functional mean curves (across participants) for the three conditions of the experiment.

implemented in Matlab. To begin, the raw data were appropriately scaled by transforming them to z scores, in order to account for differential use of the slider device. Each of the 30 observation records consisted of a vector of length 800, from data collected every 10 milliseconds for the duration of the 80-second performance (one participant had to be excluded due to equipment failure, resulting in a total n of 29). The values ranged from 0 to 127 depending on how much of the range of the slider each participant used. To account for differences in individual use of the slider, each vector was scaled by its median. That is, the median of a given vector was subtracted from each of the 800 observations in that vector. This procedure centres each participant's ratings so that a score of zero refers to the middle-most tension felt by the participant during the performance. Then the vectors were scaled so that the maximum score corresponds to a value of 100 and the minimum a value of 0, so each participant's ratings are comparable on an interval of $[0, 100]$. Note that this scaling does not equate the level of tension between two or more participants; rather it allows us to compare the maximum and minimum tensions reported by participants during their experience of the performance.

Functional objects were then created from the discrete, scaled vectors. That is, each vector of observations was represented as a smooth curve. First, 150 Order 6 B-splines were used as a basis expansion for each curve. In a sense, this represented an initial reduction of data, since the 800 observations were now summarized using 150 basis coefficients. Order 6 splines were used because they ensure that the first and second derivatives of the resulting curves would be composed of Order 5 and Order 4 B-splines, respectively and thus still be smooth (DeBoor, 2001; Shikin & Plis, 1995). Then each curve was smoothed by a small amount to take out a bit of roughness.

The raw data were previously shown in Figure 1

and the functional objects created from them are now displayed in Figure 3. A noticeable alignment can be seen at about the 30-second mark in the Audio-Only group. This sudden rise and then drop in tension was seen less clearly in the Audio + Video and Video-Only groups (not shown). It is the goal of the analysis to find the times in the performance when the three groups differ most strongly and to quantify what form these differences take. To facilitate this next goal, mean functional curves across participants were created for each of the three conditions. These mean functional curves are shown in Figure 4. Among the first things we note in a visual inspection of the resulting figure is that the *auditory-only* judgments track quite closely to the *auditory + visual* judgments throughout most of the piece. Note also that the *visual* component diverges from the *auditory* and the *auditory + visual* at certain points in the piece, notably around second numbers 18 and 34. We will return to this shortly.

Functional Principal Components Analysis

As in a multivariate dataset, a good exploratory technique for these curves is a principal components analysis. In the functional case, the principal components are curves instead of vectors; however, the interpretation is very similar: The resulting principal component functions highlight the directions in which the dataset most varies. We performed a functional principal components analysis (*fPCA*) on the entire dataset to explore the modes of variation in all the participants. The principal component functions were smoothed slightly to aid in interpretation. Then the first three principal components were rotated using the VARIMAX criterion, three having been chosen here as in multivariate analysis on the basis of a satisfactory amount of variance having explained the components being interpretable. This procedure attempts to transform the principal component curves so that

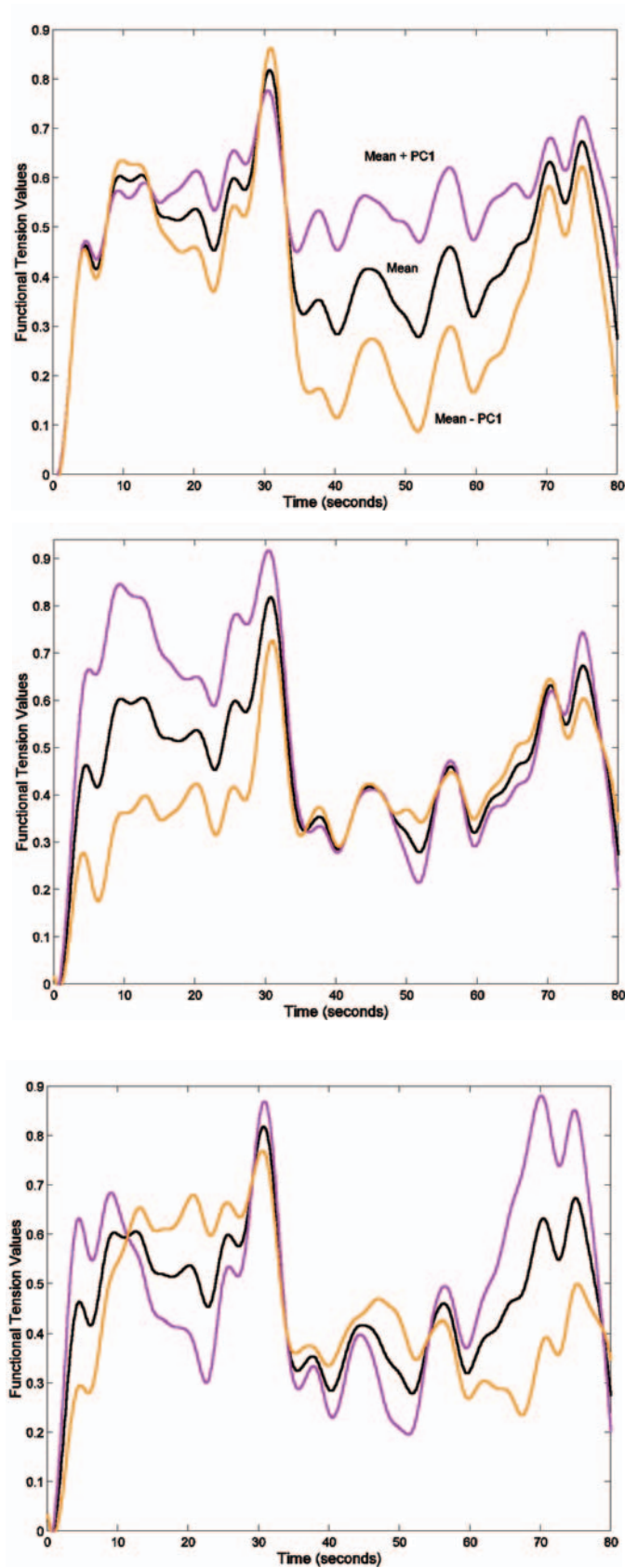


Figure 5. Effects of the rotated functional principal components (fPC) on tension judgments. In all three panels, the mean of participants' tension judgments is shown as the black line; the purple line shows the mean plus the principal component; the yellow line shows the mean minus the principal component. Top panel: The first component, accounting for 48% of the variance, indexes the degree to which a subject's judgments are expanded or compressed (amplified or attenuated) from the mean. Middle panel: The second component, accounting for 15% of the variance, indexes "start" effects, that is, the degree to which participants' judgments started out high or low. Bottom panel: The third component, accounting for 14% of the variance, indexes edge effects, that is, the degree to which the beginning and ending of the piece are seen as more or less extreme than the middle of the piece.

their values are either shrunk to zero or strongly positive and negative. This aids in the interpretation of the curves and does not affect the orthogonality of the components. The following graphical tool is helpful to visualize the effects of the principal components: The mean curve for the dataset is first plotted; then a small multiple of each principal component is added to and subtracted from this mean curve and plotted on the same axes. Thus we can more clearly see the times in the performance where the participants feel strongly differing tensions.

The effect of the first functional principal component is displayed in the top panel of Figure 5 alongside the mean of participants' judgments. The principal components in FDA are curves. We can see that the effect of the first extracted component is that of attenuation or amplification – we might label this component “compression/expansion,” referring to the strength of an individual's reactions to the underlying tension events. Participants with a high score on this component (purple line) experience an amplified effect as compared to the mean effect: a strong drop in tension at about Second 35. These participants also report a sharp increase in tension at about Second 65 after a period of low tension for 30 seconds. On the other hand, participants with a low score on this component (yellow line) report a more attenuated experience: The drop in tension at Second 30 is not as noticeable, and the tension remains high for the next 35 seconds and beyond.

The effect of the second principal component is displayed in the middle panel of Figure 5. It appears that this component identifies the “start effects” of the participants' experiences. A high score on this component indicates a participant who reports a high level of tension in the beginning of the piece but then experiences typical tension throughout the rest. A participant with a low score on this component starts with a level of tension lower than average.

The effect of the third principal component is displayed in the bottom panel of Figure 5. This appears to identify the “edge effects” of the ratings. Participants with a high score on this component will begin with a high rating of tension, immediately compensate with a lower rating, and then conclude the piece with a high rating again. Those with a low score will begin with a low rating and also conclude with a low rating.

Functional Linear Model

A linear model of the tension ratings was built from the functional data to better quantify the effects suggested by the principal components analysis. Recall that in the musical tension study, participants

were assigned to one of three conditions in a between-participants design: Audio + Video (the natural condition, in which the performance was both heard and seen), Audio Only (the sound was on but the image was turned off), and Visual Only (the sound was off but the image was turned on). The predictors were scalars: dummy variables indicating the group (or condition) membership of the participant producing each tension rating. Since the responses were functional objects, the resulting regression coefficients were also curves rather than scalars, and they allow us to determine where in the performance the experimental condition most affects the tension ratings. The coefficient curves were smoothed slightly to aid in interpretation.

The linear model can be written as

$$y_i(t) = \beta_0(t) + \beta_1(t) AR_i + \beta_2(t) VR_i + \varepsilon_i(t). \quad (10)$$

Here the dummy variables AR and VR are such that $AR_i = 0$ if the i th participant experienced the performance with the audio removed (Video Only), and 1 otherwise, and $VR_i = 0$ if the i th participant experienced the performance with the video removed (Audio Only), and 1 otherwise. So an average tension rating for the participants in the Audio-Only group would be $\beta_0(t) + \beta_2(t)$; for those in the Video-Only group it would be $\beta_0(t) + \beta_1(t)$; and for those in the Audio + Video group it would be β_0 . Notice that this linear model is built with the baseline effect being the presence of both audio and video rather than the absence of both. Furthermore, the coefficients refer to the effects of the removal of sensory inputs rather than the presence of them. This is because the condition of no audio and no video could not be measured easily, so the standard of reference must be the “natural” condition of Audio + Video, and changes to this condition measured in subtraction rather than addition.

The coefficient curves for this linear model are plotted in Figure 6. The baseline curve of the natural condition (blue line) shows a moderate drop in tension at about Second 30, as expected, with a subsequent increase in tension at Second 65 and beyond. The “video removed” coefficient (red line) shows the effect of removing the video component of the performance. As in a scalar regression coefficient, values near zero indicate a smaller effect of the predictor, and strongly positive or negative values point to a larger effect. From this we can see that, compared to the “nothing removed” (Audio + Video) condition, the “video removed” (Audio Only) group does not experience much difference in tension from the beginning of the performance *until* about Second 30. From this

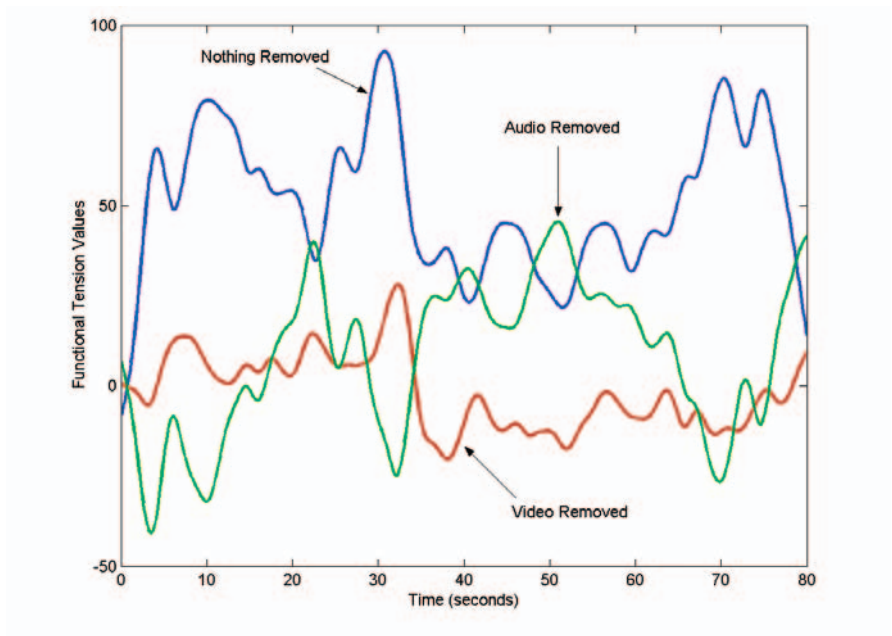


Figure 6. Coefficient curves for the functional linear model of tension. These curves allow one to visualize the effects of removing one of the sensory channels (audio or video) from the “natural” condition of audio + video.

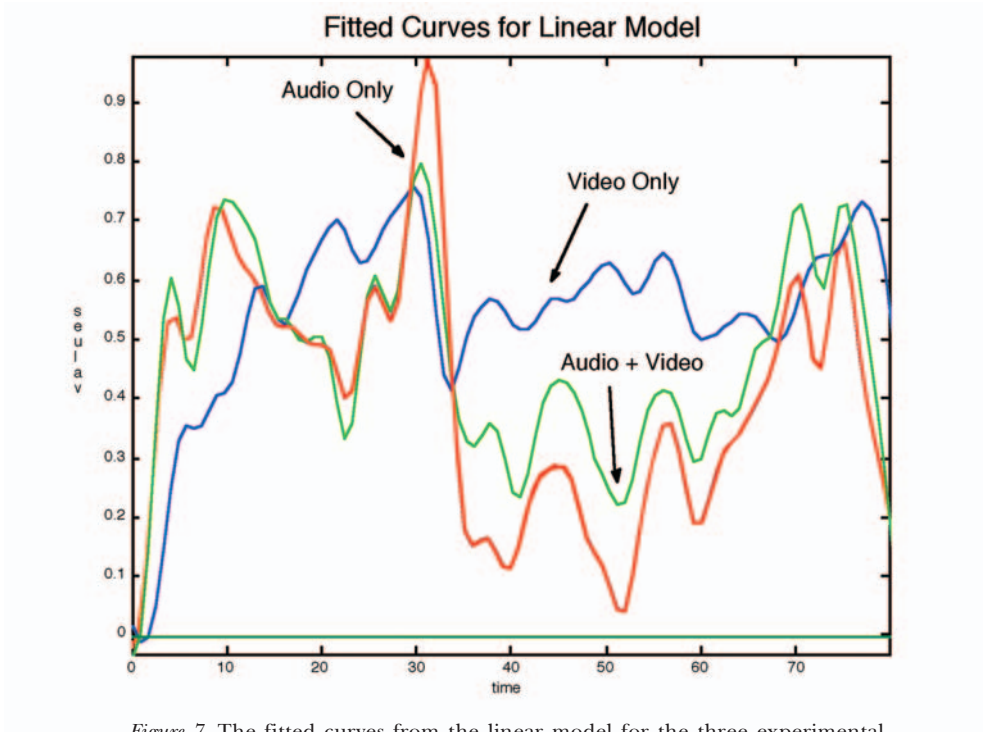


Figure 7. The fitted curves from the linear model for the three experimental groups, averaged across participants.

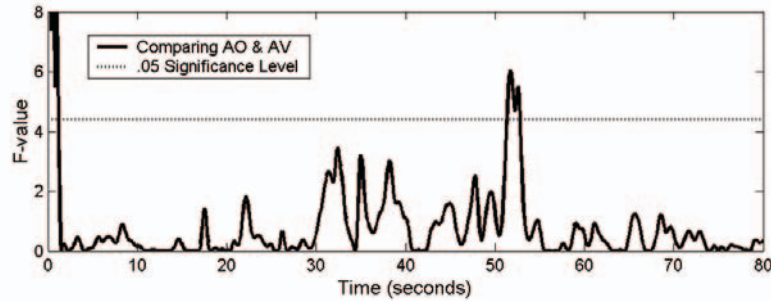


Figure 8. Comparing means for the Audio + Visual and Audio-Only groups. The dashed line represents the F-value for the hypothesis that the two group means are different with $\alpha=.05$. When the F-value rises above the dotted line, the difference between groups is significant.

point until the end of the piece, the effect of removing the video is to decrease the tension as compared to the baseline group. This effect is moderately low, meaning that the Audio-Only group does not differ greatly from the Audio + Video group throughout the piece. On the other hand, the coefficient for the “audio removed” (Visual-Only) component (green line) shows a stronger effect. Areas of greater change in tension are seen specifically at about Second 5, Second 10, Second 20, Second 30, and especially from Second 35 to Second 65. A sharp decrease in tension at Second 30, reported by participants experiencing the audio component, was not seen by participants in the Video-Only group. And their subsequent experience from Second 35 to Second 65 was not one of substantially lowered tension, as it was with the Audio-Only and Audio + Video groups.

Fitted values from the linear model for the three experimental groups are plotted in Figure 7. Here the net effect of the regression coefficient curves can be more clearly seen. The Audio-Only group (red line) showed judgments that were quite similar to the Audio + Video (green line) group for the duration of the performance, differing only in intensity at the peak at Second 30 and during the periods of lull and closure from Second 35 to Second 75.

The Video-Only group (purple line), on the other hand, shows a markedly different pattern of tension throughout the piece. Most noticeable are the attenuated peak at Second 30 and the report of average tension from Second 35 to Second 65. During this latter span of 30 seconds, the *visual* signal is evidently responsible for higher tension ratings than the audio or audio + visual. This led the experimenters to re-examine the original performance in an effort to understand why this might be so. The point at which the visual information first breaks apart most clearly from the other channels (around Second 32) coin-

cides with a long rest in the piece. The performer draws the clarinet away from his mouth, relaxes his entire body for one second or so, and then takes a deep breath in preparation for playing again. There is no evidence of this activity in the audio-only channel, but in the visual channel it is quite obvious that the performer is making preparatory gestures in anticipation of playing. This gesture serves as a cue to the viewer that something is imminent, and may underlie the ratings of increased tension at this point. Why the ratings in the visual condition stay uniquely high for the next 30 seconds – or why the audio + video participants did not respond similarly to this visual cue – remain topics for further investigation.

Functional Analysis of Variance (fANOVA)

A functional analysis of variance (*fANOVA*) can be performed on the regression curves derived in the previous section. Here, for the sake of simplicity, we have chosen to conduct a one-way *fANOVA* with presentation condition as the between participants factor, and we have analyzed two conditions: the audio-video (AV) and the audio only (AO). H_0 is that these two curves are the same, H_1 is that they are different. Our research question asks if the addition of visual information to the auditory information makes a significant impact on the experience of participants, and if so, how much? Perhaps most interestingly, we are now in a position to ask *when* in the performance these differences are significant.

Referring to Figure 8, the y-axis shows values of the *F* function, and the dashed line represents the .05 level of significance, at $F = 4.4$ (for $df = 2, 18$). FDA tools calculated this *F*-value at every point in time; when the *F*-value is above 4.4, the difference between the AV and the AO groups is significant at .05. (At present, such tools are rather primitive and do not take into account family-wise error rates, nor do they have an

easy means of determining power. These issues are under development.) The presence or absence of visual stimuli differentiates the experience for the two subject groups, so the visual component (or its interaction with the sound) is assumed responsible for any differences between the curves. We see clearly that there exists a significant section that lasts from 51.3 seconds to 53.8 seconds. The researchers closely examined the frames of the video corresponding to this section and discovered that the music and the performer's gestures are in some sense at odds with one another, which may account for the statistically significant difference between AV and AO here. Musically, the piece becomes very soft. Visually, the performer takes on an intense expression in his face with his eyebrows raised, and his body moves in a rising motion even beyond the sounding of notes, during rests. All of this visual activity contrasts with the sonic content, which conveys a sense of total calm and quiet. The performer's body movements and facial expressions clearly indicate that he had a highly emotional experience during this portion of the performance. This high level of emotion was conveyed to the AV participants via the visual modality. The affective response of the AV group, as indicated by the tension judgment, maintained a significantly greater magnitude than that of the AO group because of the visual content in the performance. This is a subset of the period of time during which participants in the visual-only condition were also showing high amounts of tension.

Summary of Findings

The auditory and visual components of this particular performance conveyed some of the same experience of tension, during particular portions of the piece. It has long been suggested in theories of art and aesthetics that the *gestures* made by an artist reflect *emotional intent*, even when the gestures themselves are not the primary object of evaluation by the perceiver. For example, in Japanese ink paintings, the gestures used by the painter and the resulting brush strokes are believed to convey the emotional state of the artist at the precise moment of the gestures (Davey, 1999). The data here provide evidence that the gestures and movements of a performing musician – even in the absence of sound – are conveying some of the same information as the sound itself in many cases. In other cases, as we have seen, the visual channel is providing complementary and different information than the audio channel. Much remains to be studied in this domain. We have included this example both because of its novelty and because it illustrates that certain interpretations of a dataset are made easier with FDA.

Other Examples

There are many other examples of functional data to be found in psychology and many of the techniques just described would be applicable in these cases as well. Many studies in psychology, electrical engineering, and linguistics concern temporal synchrony/asynchrony. For example, Cummins (2003) studied speech synchronization: Native speakers of English were asked to simultaneously read an unfamiliar passage of text, and the degree to which their utterances were temporally synchronized was studied. Of interest was whether or not the participants improved with practice. Replications existed at the level of participants ($n = 27$) and at the level of trials (each participant read five text excerpts). For each paired reading ($n = 27 \times 5 = 135$), the median asynchrony, in milliseconds, was computed at three temporal positions: the onset, medial, and final vowel onsets. FDA would permit the researcher to construct a synchrony curve for each participant pair – or for the participants all taken together – and to analyze the first versus last trial, or the five trials together, without having to conduct tests at only three time points. FDA thus allows the researcher to use all of the information available – not just those at certain time points. This could reveal any latent information about when during the trials participants were maximally or minimally synchronized, and an analysis of derivatives could reveal information about the manner in which participants “sped up” or “slowed down” in order to achieve synchrony. These micro-variations in time synchrony could inform cognitive theories of time-keeping and central clock mechanisms.

In studies of motor planning and coordination, tapping tasks are often employed (Rosenbaum, Kenny, & Derr, 1983; Schulze, Lüders, & Jäncke, 2002). Participants are asked to tap their fingers either to a presented sequence, or to their own “preferred tapping rate.” Again, FDA allows the researcher to move beyond summary statistics that collapse over the time period of the experiment, to more closely analyze effects *as a function of time*, as well as the derivatives of any timing and motor responses obtained.

Evoked Response Potentials (ERPs) are collected by placing electrodes on the surface of the scalp, and electrical activity is measured in response to various stimuli or tasks the subject is asked to complete. In such studies, the ERP signal is typically collected from 32, 64 or 128 channels distributed around the skull, and it is of interest to understand how the ERP signal varies across time and across sites. Somewhat similarly, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) track the blood-oxygenated level-dependent signal in the brain, and

reveal brain regions and neural networks that are activated during particular cognitive tasks. These studies produce millions of datapoints every few seconds, and the points are correlated both in space (a group of neurons is likely to be active if an adjacent group of neurons is) and in time (neurons active at any time t are likely to still be active at $t + 1$). These data thus consist of highly intercorrelated observations, and are most often analyzed using statistical methods that attempt to correct for multiple comparisons and such intercorrelations (e.g. Friston et al., 1995; <http://www.fil.ion.ucl.ac.uk/spm/>; <http://www.mrc-cbu.cam.ac.uk/Imaging/Common/>). FDA techniques are beginning to be applied to neuroimaging data (Beckman & Smith, 2004; Valdes-Sosa, 2004; Viviani, Grön, & Spitzer, 2003).

Other recent applications of FDA include the analysis of esophageal bolus flow (Stier, Stein, Schwaiger, & Heidecke, 2004) and analysis of gene expression arrays (Barra, 2004).

Conclusions

We described a number of places in the behavioural and life sciences research in which functional data may be found, and the tools we have outlined here could potentially reveal relations not as easily seen with conventional statistical approaches. In many cases, FDA will allow the researcher to ask questions that either could not be asked, or would be computationally cumbersome to ask, using traditional statistical methods. In this respect, FDA represents both novelty in analysis and parsimony in execution. Moreover, functional models facilitate a visual representation of the data, and thus provide explanatory power.

We illustrated key aspects of FDA using real data from a study of musical emotion. Note that in this example, as in most functional datasets, several key features were present:

- The data were drawn from a continuous measure (position of a slider that theoretically varied continuously both over its range and over time);
- We can assume that the process generating the data is a smooth, continuous one;
- The dataset contained replications (10 participants in each of three conditions);
- The curves are differentiable, and the derivatives are a potential source of additional information about the nature of the data; and
- Each functional data curve constitutes an observation (rather than the individual points that make up the curve being considered observations).

We showed that an advantage of employing FDA for datasets such as this one is that we can exploit these features to reveal information not otherwise attainable in the data. In particular, amplitude and phase variability in the curves can be examined or eliminated (depending on the researcher's hypothesis), curve registration allows us to view and analyze the underlying process of interest, and derivatives of the curves can be analyzed to reveal dynamic aspects of the process. In addition, FDA contains methods for constructing correlations and Principal Components that are themselves curves, and thus allow us to track *when* in the time series (or *where* in a spatial series) any significant differences between and among curves exist. In constructing functional F -tests, we can avoid the problems of multiple comparisons inherent in treating each *point* on the curve as an observation, by creating F -function curves that relate directly to the dynamic character of the data.

FDA comprises a set of rapidly evolving tools. FDA comprises a set of rapidly evolving tools. A good deal of this research energy goes into refining existing methods such as the functional linear model and principal components analysis, but we tend to see two problems specific to FDA that are attracting particular attention. The first of these is the *simultaneous* modeling of amplitude and phase variation, as opposed to the procedure described in this paper of using registration methods to remove phase variation before modeling amplitude variation. This integrated approach seems important because amplitude and phase can interact in interesting ways, and also because independent variables that affect amplitude variation can also affect phase as well. This is obviously the case for the effect of gender on growth curves, for example. A second general trend is toward using *dynamic* models, often expressed in terms of systems of differential or difference equations, to study rates of change directly. A good sample of current work of this nature can be found in Walls and Schafer (2006). We provided links to a web site which is constantly updated so that the reader can access software tools to perform his or her own functional data analyses.

This paper was supported by grants from NSERC to DJL and JOR.

Correspondence concerning this article should be addressed to Daniel J. Levitin, Department of Psychology, McGill University, 1205 Avenue Penfield, Montréal, Québec H3A 1B1 (Phone: (514) 398-8263; Fax: (514) 398-4896; E-mail: dlevitin@psych.mcgill.ca).

Résumé

Les psychologues et spécialistes du comportement collectent de plus en plus des données tirées des processus sous-jacents continus. Dans le présent article, nous décrivons un ensemble de méthodes quantitatives, l'analyse de données fonctionnelle, qui peut répondre à de nombreuses questions auxquelles les approches statistiques traditionnelles ne peuvent répondre. Ces méthodes sont utiles pour l'analyse de données communes en psychologie expérimentale, notamment les données visant des séries temporelles, les mesures répétées et les données s'échelonnant dans le temps ou l'espace comme c'est le cas pour les expériences en neuro-imagerie. Le principal avantage de l'analyse de données fonctionnelle est qu'elle permet au chercheur d'examiner, lors de l'analyse de séries temporelles, quand des différences peuvent-elles exister entre deux ou plusieurs ensembles. Notre discussion traite des corrélations fonctionnelles, des composantes principales, des dérivés des courbes fonctionnelles et de l'analyse des modèles de variances.

References

- Barra, V. (2004). Analysis of gene expression data using functional principal components. *Computer Methods and Programs in Biomedicine*, 75, 1-9.
- Beckmann, C. F., & Smith, S. A. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23, 137-152.
- Boker, S. M., Xu, M., Rotondo, J. L., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, 7:1, 338-355.
- Box, J. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. Englewood Cliffs, NJ: Prentice-Hall.
- Cleveland, W. S. (1993). *Visualizing data*. Murray Hill, NJ: AT&T Bell Laboratories; Summit, NJ: Hobart Press.
- Cudeck, R. (1996). Mixed effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, 31:3, 371-403.
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31, 139-148.
- Davey, H. E. (1999). *Brush meditation: A Japanese way to mind & body harmony*. St. Paul, MN: Stone Bridge Press.
- De Boor, C. (2001). *A practical guide to splines, revised edition*. Berlin: Springer-Verlag.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
- Fernández de Castro, B., Guillas, S., & Gonzáles Manteiga, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, 47(2), 212-222.
- Freedman, D. (1984). On bootstrapping two-stage least squares estimates in regression problems. *Annals of Statistics*, 12, 827-842.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2, 189-210.
- Gauss, C. F. (1809/1963). *Theory of the motion of the heavenly bodies moving about the sun in conic sections*. New York: Dover Publications.
- Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. London: Chapman and Hall.
- Guo, W. (2002). Functional mixed effects models. *Biometrics*, 58, 121-128.
- Guo, W. (2003). Functional data analysis in longitudinal settings using smoothing splines. *Statistical Methods in Medical Research*, 13, 1-14.
- Hendry, D., & Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, 21, 1-42.
- Huitema, B. E., & McKean, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods*, 3(1), 104-116.
- Kneip, A., Li, X., MacGibbon, B., & Ramsay, J. (2000). Curve registration by local regression. *Canadian Journal of Statistics*, 28, 19-29.
- Koenker, R., & Mizera, I. (2004). Penalized triograms: Total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society Series B*, 66, Part 1, 145-163.
- Krumhansl, C. K. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- McAdams, S., Winsberg, S., Donnadieu, S., & De Soeta, G. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177-192.
- McKnight, S. D., McKean, J. W., & Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, 5(1), 87-101.
- Nuzzo, R. L. (2002). *Stochastic models for lipid biochemistry*. Unpublished doctoral dissertation, Stanford University, Palo Alto, CA. (DAI-B, 64/03, UMI #AAT 3085218).
- Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of*

- Electrical Engineers*, 47, 617-644.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg, (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397-414). New York: Sijthoff, Leiden.
- Ramsay, J. O. (1995). A similarity-based smoothing approach to nondimensional item analysis. *Psychometrika*, 60, 323-339.
- 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. (2000). Functional components of variations in handwriting. *Journal of the American Statistical Association*, 95, 9-15.
- Ramsay, J. O. (2002). Multilevel modeling of longitudinal and functional data. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ramsay, J., & Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society, Series B*, 53, 539-572.
- Ramsay, J. O., & Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B*, 60, 351-363.
- Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. New York: Springer-Verlag.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis, second edition*. New York: Springer-Verlag.
- Rosenbaum, D. A., Kenny, S., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 86-102.
- Schulze, K., Lüders, E., & Jäncke, L. (2002). Intermanual transfer in a simple motor task. *Cortex*, 38, 805-815.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37, 10-21.
- Shao, X. M., & Chen, P. X. (1987). Normalized auto- and cross-covariance functions for neuronal spike train analysis. *International Journal of Neuroscience*, 34, 85-95.
- Shepard, R. N. (1972). Psychological representation of speech sounds. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill.
- Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of music* (pp. 343-390). San Diego, CA: Academic Press.
- Shikin, E. V., & Plis, A. I. (1995). *Handbook on splines for the user*. Boca Raton, FL: CRC Press.
- Simonoff, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.
- Sinnott, J. M., Brown, C. H., Malik, W. T., & Kressley, R. A. (1997). A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Perception & Psychophysics*, 59, 1214-1224.
- Stier, A. W., Stein, H. J., Schwaiger, M., & Heidecke, C. D. (2004). Modeling of esophageal bolus flow by functional data analysis of scintigrams. *Diseases of the Esophagus*, 17, 51-57.
- Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Unser, M. (1999). Splines: A perfect fit for signal and image processing. *IEEE Signal Processing*, 16, 22-38.
- Valdes-Sosa, P. A. (2004). Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics*, 2, 239-250.
- Vidakovic, B., & Ruggeri, F. (2001). BAMS method: Theory and simulations. *The Indian Journal of Statistics. Special Issue on Wavelet Methods*, 63, 234-249.
- Vines, B. W., Nuzzo, R. L., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Crossmodal interactions in the perception of musical affect and structure: The impact of seeing a clarinetist perform. *Cognition*, 101, 80-113.
- Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics and functional data analysis techniques. *Music Perception*, 23:2, 137-152.
- Viviani, R., Grön, G., & Spitzer, M. (June 2003). *Functional principal component analysis as an explorative tool for fMRI data*. Poster #970 presented at the Human Brain Mapping conference, New York.
- Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for intensive longitudinal data*. Oxford: Oxford University Press.