In Speech Planning and Dynamics, p. 239-277. S. Fuchs, M. Weirich, D. Pape & P. Perrier, eds. (2012) Frankfurt: Peter Lang.

A survey of methods for the analysis of the temporal evolution of speech articulator trajectories

LEONARDO LANCIA
MARK TIEDE

Abstract: In this chapter we will survey methods for comparison and description of sampled movement data, with focus on techniques able to capture the characteristics of the temporal evolution of the signals under study. Functional data analysis, the cross wavelet transform and cross recurrence analysis will be introduced and applied to the analysis of speech movement data.

1 Introduction

Speech activity can be observed through a variety of transduced signals: these can be acoustic signals, aerodynamic signals, recordings of the movements of the speech articulators, etc. Such signals evolve over time in a generally continuous fashion, describing trajectories that are characteristic of both the linguistic content of the utterances produced and of non linguistic aspects such as the emotional or health state of the speaker, the communicative context, and so forth. In speech production research, it is common practice to focus on particular events present in these signals such as peaks, valleys or discontinuities. As a consequence, speech is often quantified through measures anchored to particular time points or averaged over portions of the signals where no discontinuity is observed. In such an approach, the critical points of analysis must be defined before the analysis takes place, with many evident drawbacks: information concerning the trajectories themselves is lost; the results are strongly dependent on the theoretical assumptions

that lead to the choice and the location of the events on the signal; and a large amount of time-consuming manual labeling is generally required. In recent years, new methods have been proposed for characterizing the evolution over time of multivariate time series and which offer viable alternatives for studying the signals recorded in speech production experiments. The purpose of this chapter is to contribute to the diffusion of these methods with a tutorial introduction focused on the analysis of the motion of oral articulators.

1.1 The Labial Coronal Effect

The application of these techniques will be illustrated using the movements of three articulators observed by electromagnetic articulometry (EMA): the jaw (as characterized by the movement of the lower incisors), the tip of the tongue (henceforth TTip) and the lower lip (henceforth LLip). The data were collected using a speeded repetition paradigm to replicate the so-called *labial coronal effect* observed by Rochet-Capellan and Schwartz (2007). In several languages of the world, disyllabic CVCV sequences in which the two consonants are a labial and a coronal are found more often in labial-coronal order than in coronal-labial order. A hypothesis defended by Rochet-Capellan and Schwartz is that this preference is observed because the labial-coronal order (henceforth La-Co order) is inherently more stable than the reverse order (Co-La). To test their hypothesis, Rochet-Capellan and Schwartz had French speakers repeat Co-La or La-Co disyllabic nonsense words (eg. /pata/, /tapa/). Nonsense words were uninterruptedly repeated during trials of 10 seconds. In the first half of each trial, speakers had to gradually increase their speech rate and in the second half the speech rate was gradually decreased. Regardless of the order of the consonants in the intended utterance, at fast speech rates speakers tended to produce the sequences with an underlying La-Co order.

An important observation made by Rochet-Capellan and Schwartz concerns the evolution of the jaw's behavior during a sequence of utterances. At slow speech rate, two upward movements of the jaw are observed per disyllable. This is expected, because in the presence of an oral constriction, the jaw is generally raised (Lindblom, 1983; Keating et al., 1994) However, at fast speech rates, the upward movements tend to be

reduced to one per utterance. In a replication of the Rochet-Capellan & Schwartz study applied to German, Fuchs et al. (2009) emphasize the presence of several instances in which at high speech rates the jaw tends to be frozen, and its vertical displacement is minimal. This behavior was also observed by Rochet-Capellan and Schwartz, who explicitly discarded from the analysis the utterances produced in this way. On the basis of this behavior Fuchs and colleagues were led to emphasize the role of the jaw in the production of the Labial Coronal effect.
Subsequently Lancia and Fuchs (2011) formulated a revised hypothesis in which this effect is promoted by the interaction of the two following constraints operating on different levels: 1) As speech rate increases, it becomes increasingly difficult to produce two jaw oscillations per utterance, due to both biomechanical and cognitive factors. Because the jaw is the most massive articulator, preserving its oscillatory frequency as the double of the oscillatory frequency of the other, less massive articulators during several sequences in which dozens of repetitions are produced becomes untenable. In addition, when the jaw produces two oscillations per utterance, the ratio between the oscillatory frequency of this articulator and the oscillatory frequency of the other two articulators is 2:1. It is known that this kind of coordination pattern is generally harder to maintain than the 1:1 frequency ratio which characterizes the coordination of the articulators when one jaw cycle per syllable is produced (Pouplier, 2007; Goldstein et al., 2007). 2) To produce an audible coronal gesture, the jaw must achieve a high position; at the same time, the constraints on the height of the jaw during the labial constriction seem to be weaker (Lindblom, 1983; Keating et al., 1994; Hertrich and Ackermann, 2000; Koenig et al., 2003; Mooshammer et al., 2007). According to Lancia and Fuchs (2011), the joint effect of these constraints produces a weakening of the jaw cycle which is synchronous with the movement of the LLip. The reduction of this cycle of jaw movement produces a reduction of the vowel following the labial consonant and a reduction of the interval between this consonant and the following coronal. For this reason the two consonants tend to be perceived as grouped as if they were produced with a LC order. However, in order to test this revised hypothesis, three tasks must be performed. First, the evolution from two jaw cycles per utterance to one jaw cycle per utterance has to be quantified. Second, relative coordination among the articulators has

to be determined, together with how this changes as speech rate is increased. Finally, the variability of fast speech rate utterances produced with an underlying La-Co must be compared to those produced with an underlying Co-La order.

1.2 Plan of the chapter

The goal here is not to test the hypotheses described in the preceding section, but rather to explain and illustrate the analysis methods adopted to accomplish the three tasks described above. In the following sections, we will first introduce some basic terms. We will then illustrate how functional data analysis can be used to characterize the evolution of the jaw's behavior as speech rate is increased. In order to relate the behavior of the jaw to the other articulators, we will use the cross-wavelet transform to derive their relative phase relationships and show the evolution of these variables as speech rate increases. We will also introduce recurrence quantification analysis as a method for assessing the variability of motion patterns corresponding to different consonant orders (Co-La or La-Co). In the last section of this chapter we will briefly discuss choices confronting the researcher for the correct application of these methods. In the appendix, the reader is directed to essential literature describing these techniques and to the internet resources providing software and examples.

1.3 Basic terminology

The data studied here have been obtained using electromagnetic articulometry (EMA), in which signals transduced in sensors within an oscillating electromagnetic field are used to track their positions and orientation in three dimensions. By gluing sensors to accessible parts of a speaker's vocal tract anatomy (anterior tongue, jaw, lips) and correcting for head movement using additional reference sensors, the simultaneous and overlapping movements of the speech articulators relative to the hard structure of the palate can be observed using this technique. At time t the position of each of these sensors can be characterized by three spatial coordinates corresponding to its 3D offset from a common origin (typically the speaker's upper incisors). The set of values that

constitute the spatial configuration of the ensemble of the articulators is a coordinate vector. When dealing with position data, the values correspond to spatial coordinates, but when dealing with other kinds of data (eg. pressure, flow, etc.) each vector describes the state of the observed signals at a given moment in time and it can therefore be called a **state vector**. A **multivariate time series** is a sequence of state vectors describing the simultaneous evolution of several processes, each one projected onto a different dimension of the state vectors. In the case of speech movements, a given dimension of the state vectors (and, by extension, of the whole time series) corresponds to the evolution over time of the position of one particular articulator over one of the three spatial dimensions. It is often the case that one or more spatial dimensions are dropped from analysis, on the assumption that movement along those dimensions is not relevant for the phenomenon under study. The sequences of values on the different dimensions, when plotted separately, constitute **curves** on a plane which have time on the abscissa and the displacement mapped to the ordinate. Particular shapes determined by the evolution over time of one curve or by the joint evolution of several curves are defined **spatio-temporal patterns**. It is often the case that we consider a one dimensional signal as the combination of simpler curves which are called **components**. This approach can of course be extended to multidimensional signals. When analyzing a roughly periodic signal (a signal which is repeated in a more or less cyclical fashion over time), it is common to decompose it as the sum of basic oscillations or building-blocks occurring at different frequencies. In Fourier analysis, the components are sinusoidal functions of time multiplied by appropriate constants. Other analysis methods adopt different building-blocks, as shown in the two following sections.

2 Part I: Functional data analysis (FDA)

2.1 Introduction

Rochet-Cappellan and Schwartz observed that when speakers repeat CVCV disyllables at a moderate speech rate their jaw produces two oscillations. The vertical position of the jaw reaches its maximum at con-

sonant closure, and its minimum corresponds to vowel production. At high speech rates however the jaw often makes only one oscillation per utterance. In the first part of this chapter we will analyze the trajectories of the jaw from 20 /pata/ utterances produced in the same trial sequence.
In order to compare the trajectories and to understand how they depend on speech rate, we follow Ramsay and Silverman (1997) and consider each trajectory X as the instance of a smooth function of time.

$$X \approx F(t) \tag{1}$$

Such a function can be approximated by a linear combination of primitive functions of time. The behavior of the primitive (or basis) functions is well known and can be adapted to approximate the behavior of the raw data. This is done by tuning the multiplicative coefficients of the basis functions.

$$X = \sum_{i=1}^{n} c_i f_i(t) \tag{2}$$

with $t = \{1, \ldots, m\}$. In other words, the time series X, of length m, can be described as the sum of n functions f multiplied by n coefficients c_i.
The term *basis function* is used to refer to one of the functions f, while the term *mother function* refers to the basic shape of the basis functions. n basis functions and n coefficient values constitute a functional observation, i.e., the functional model of an observed trajectory. In a way, a functional observation is a smoothed version of the original data, and indeed smoothness of trajectories is a prerequisite for that computational method. The first step in FDA is thus the transformation of our original observations into *functional observations*.
The second step is related to the alignment of the trajectories. Since the lengths of any two utterances are rarely equal, if we want to compare them we need to know which features of the first utterance corresponds to which features of the second. This mapping is called *registration*.
Once the second step has been completed, we have a set of functional observations that can be compared to determine the amount of variation in different portions of the trajectories, or that can be submitted to more sophisticated analyses. In this chapter we will use functional principal component analysis, an extension of classical PCA. PCA is commonly used to characterize the extent of variation in multi-dimensional

data. As it will be shown in section 2.2.3, with the functional extension of this technique, we can distinguish temporal patterns which maximally change over the curves. In the example that follows we will use this strategy to observe that when the speech rate increases, the jaw movement changes from a repeated pattern consisting of two jaw cycles per utterance toward a pattern with only one jaw cycle.

2.2 Method

2.2.1 First step: Basis expansion

This operation is performed in two steps: 1) the definition of the building-block or basis function and 2) the computation of the coefficients to fit the basis function to the observed sequences.
In cyclic speech, where vertical displacement of the articulators has mostly a sinusoidal shape, it is appropriate to model the observed trajectories with combinations of sine and cosine functions at different oscillation frequencies, adopting a Fourier basis set. With that choice, the parameters needed to define the set of basis functions are their number (i.e. the number of sine components plus the number of cosine components plus 1 for a constant component) and the period of the slowest sinusoid in the set (by default equal to the duration of the signal). We used a Fourier basis of the 11th order with one constant coefficient and 10 sinusoidal coefficients:

$$F(t) = C_1 + C_2 \sin(\omega t) + C_3 \cos(\omega t) + C_4 \sin(2\omega t) + C_5 \cos(2\omega t) + \ldots + C_{10} \sin(10\omega t) + C_{11} \cos(10\omega t) \quad (3)$$

where ω is the lowest period of oscillation, equal to the duration of the utterance. When the signal is not periodic, a basis of B-splines is the usual choice. With B-splines, the signal is divided into a finite number of subsequences and each subsequence is approximated by a polynomial. Often cubic B-splines are used; the basis is built combining polynomials of the fourth order. (In the notation conventions adopted by Ramsey and Silverman the order of the polynomials includes the term of degree 0, so that it is equal to the degree of the polynomial plus one.). Thus the basic polynomial has the form:

$$P(t) = C_1 + C_2 t + C_3 t^2 + C_4 t^3 \quad (4)$$

The entire trajectory is built summing several polynomials of the same degree but with different values for the coefficients. In B-splines each polynomial is different from 0 only over a limited and uninterrupted (i.e. compact) portion of time. The trajectory is thus split into segments modeled by different polynomials. With B-splines the parameters that we have to define are the number of basis functions used and the order of the polynomials. The total number of basis functions used is equal to the number of internal segment boundaries or "break-points" plus the order of the spline (cubic splines being polynomials of order 4). Thus if we want to divide our trajectories into 15 sections, we will need 19 basis functions. The number of internal break points determines the sensitivity of the B-splines to local features of the trajectory and the 4th order polynomials assure that the modeled curve is smooth (its second derivative is a continuous function). The coefficients are obtained through an optimization algorithm which, constrained by a roughness penalty, proceeds by minimizing the second derivative of the smoothed curves (Ramsay and Silverman, 1997).

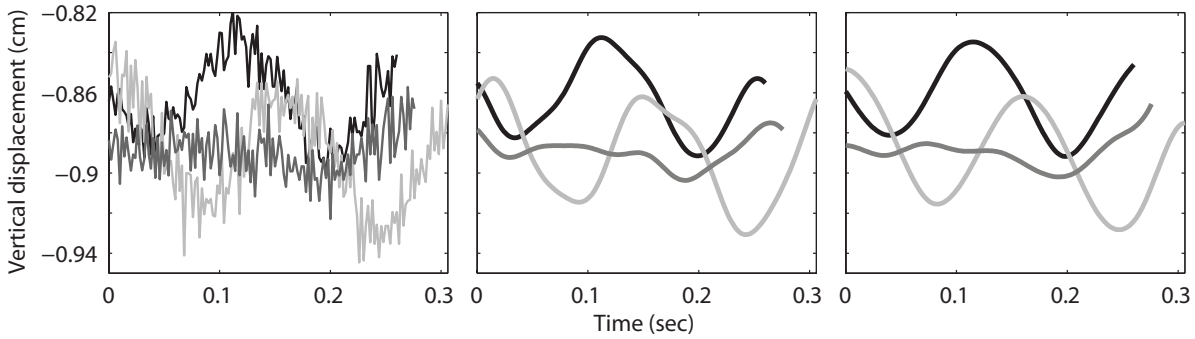

Figure 1: This picture shows the oscillations of the vertical position of the jaw corresponding to three /pata/ utterances (in different shades of gray). On the leftmost panel, the original data are shown; on the middle and rightmost panels we can observe their functional representations obtained with Fourier basis and cubic Bspline basis.

While in the middle panel of Figure 1 we see that a Fourier basis of order 11 constitutes a satisfactory choice for the raw data in the left panel, it is unfortunately easy to obtain results less consistent with the original data. The B-spline basis has its weakness in the fit at the edges of the trajectories, because in these regions a smaller number of data points are available. This problem does not affect the Fourier basis because, since the trajectories are assumed to be periodic, the missing data points at one edge can be extracted from the opposite edge. However the Fourier

basis is less flexible than the B-spline approach because it is bounded to cyclic behavior; moreover it is less sensible to the local transient behavior of the trajectories, which thus tend to disappear in the smoothed curves. It follows that, in order to apply FDA to a set of trajectories, it is necessary to check for the fit of each functional observation to the corresponding original trajectory; this can be quantified by thresholding excessive standard deviation between original and fitted data.

2.2.2 Second step: Registration

Once the functional observations have been created, their time scales can easily be normalized and the trajectories can thus vary over the same interval. In this way, the initial and terminal events correspond. However this necessary preliminary step does not assure that the portions which are equivalent across the trajectories are aligned; they are aligned only if the differences in time scale among the trajectories are uniform. In order to align the events (peaks and valleys) present in all the trajectories, the following procedure (following Lucero et al., 1997) is adopted. The trajectories are computed over the same time interval (that by convention we set from 0 to 1) and a first mean function is obtained by averaging the corresponding data points across the trajectories[1].
Next, for each functional observation, a warping function is defined as a cubic B-spline. A warping function is a transformation of the time scale. Given a functional observation, the warping function defines, for each of its points, the horizontal shift needed to align that point to the corresponding point of the mean trajectory.
The warping function coefficients are then optimized to minimize the distance between the functional observations and the mean trajectory. Applying the warping functions to their corresponding functional observations, the trajectories are aligned. This procedure may be repeated to refine the results (cf. Lucero et al., 1997; Lucero and Koenig, 2000;

1 In order to favor a correct alignment, the original curves can be centered, subtracting the mean from each curve, and scaled through division by the standard deviation (or, more appropriately when comparing time series of different length, through division by their maximum value, cf. Wang and Gasser, 1997; Lucero and Koenig, 2000). Once the aligned curves have been obtained, the initial scale and origin of the axis can be recovered by adding to each curve its initial mean and multiplying by the scale factor used to normalize the curve.

Lucero, 2005, for applications to speech movements). A new mean trajectory is obtained by averaging the aligned values of the functional observations and the warping functions are then recalculated to improve the match to the new mean trajectory.
Once the curves are aligned to the average curve (cf. Figure 2) it is easy to determine the similarity between two curves by applying a Euclidean distance metric. In the same fashion an averaged measure of variability is given by the sum of the standard deviations at different relative times divided by the number of time points, equivalent to the index of spatio-temporal variability (STI) introduced by Smith et al. (1995). However, as a consequence of the nonlinear time-warping, variation in the aligned data is primarily due to amplitude differences between trials, while the individual warping functions show variation in phase. In other words, this method, unlike Smith's linear STI, crucially allows for the separate quantification of amplitude and phase differences.

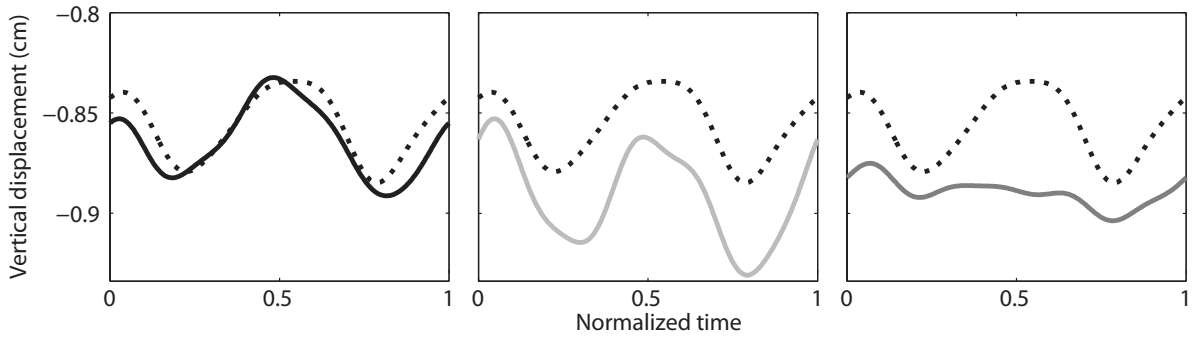

Figure 2: Aligned versions of the three curves plotted in Figure 1. The curves represent the movement of the jaw during the production of the utterance /pata/ (smoothed using a Fourier basis of the 11th order). In each panel, one aligned curve (continuous line) is plotted together with the curve obtained averaging all the 18 aligned curves belonging to the set considered (dotted line).

2.2.3 Third step: Functional principal component analysis

Given a multivariate set of data, classical PCA finds the orientation in that space over which the data vary most. We may think of the functional principal components as those basic curves whose linear combinations give the best approximation of the original set of functional observations. Therefore each curve of the set can be approximated by a linear combination of the principal components. This is a concise way to

express that each observed curve can be obtained by summing the principal components, each one multiplied by a different factor. We can thus think of a given principal component as a basic shape which is present with different intensities in the curves analyzed (cf. Gubian et al., 2010; Aston et al., 2010).
The first principal component is that basic trajectory whose variation can explain the highest percentage of the variability observed across the curves. This ranking is extended to the other components which are ordered by the percentage of total variance explained. The effect of a principal component in one particular curve is indicated by the multiplicative factor (or loading) used to approximate that curve. In this section we will focus on the interpretation of principal components but we won't address the issues raised in their extraction and by the computation of their loadings.
Before illustrating an application to real data, it is instructive to understand the meaning of principal components on a set of synthetic curves for which we know the characteristics. As an example we can consider the four curves in the right panel of Figure 3. These come from a set of 100 curves obtained from the curves $S1(t)$ and $S2(t)$ represented respectively by the continuous and the dashed lines in the left panel. These were multiplied by different constant factors (B_i and C_j respectively) and added to a sequence of random values extracted from a uniform distribution between 0 and 0.1.

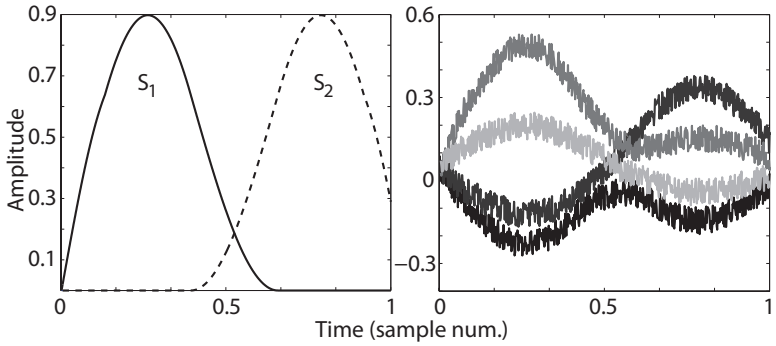

Figure 3: Left panel: basic shapes used to generate the curves in the right panel. Right panel: curves generated by combining the basic shapes in the left panel with uniform noise (see text).

The multiplicative factors B_i and C_j vary from one curve to the other and determine the amplitude of the prototypical shapes in each of the

resulting curves. The values for the factors are generated at random, but the factors C_j are multiplied by 0.8. Therefore the average effect of $S2(t)$ on the resulting curves is weaker than the average effect of $S1(t)$. In order to apply FPCA, we need to determine the number of principal components to extract. In our case, we can start with an arbitrarily high number of components, but only two of them account for non negligible proportions of the total variance observed across the curves (as shown in panel a of Figure 4). This is consistent with the fact that all of the 100 curves are obtained by combinations of precisely two basic shapes.

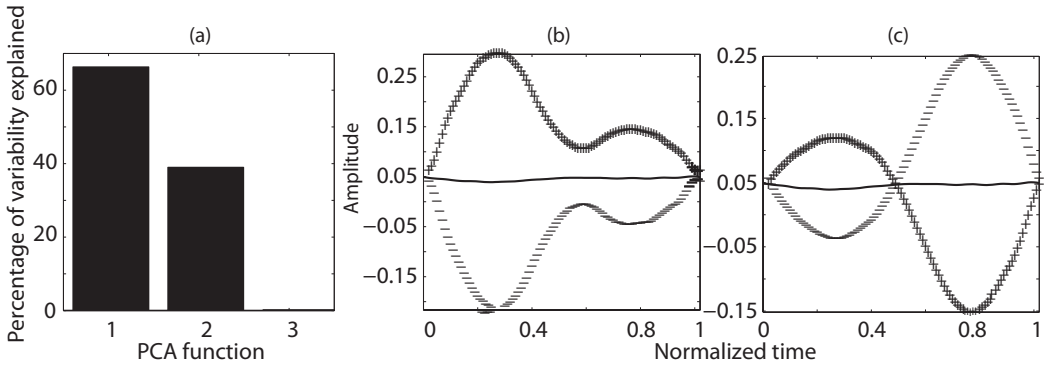

Figure 4: Panel a: Variability accounted by the first two principal components. Panels b and c: effects of the first and second principal components on the mean curve (see text).

Panels b and c from Figure 4 show the effect of the two principal components on the average curve (this is represented in both the panels by a continuous line). The two lines composed by plus signs represent the changes in the mean trajectory which are obtained when the components are multiplied by a positive factor and added to the average trajectory. The lines composed by the minus signs are obtained when the components are multiplied by a negative factor. We can see that each of the two principal components introduces two peaks (or two valleys, if combined with a negative factor); however, when adding to the average curve one of the two shapes in the left panel of Figure 3, only one peak (or one valley) is expected. We can deduce that each component accounts for variability coming from both the basic shapes. This happens because there is an infinite number of pairs of curves, different from the ones in the left panel of Figure 3, whose linear combinations can approximate the observed curves equally well. In other words, given some number of principal components, their uniqueness is not guaran-

teed. However, no better approximation can be obtained by linear combinations of the same number of curves. Using the Varimax algorithm Kaiser (1958) it is possible to select among the possible sets of principal components for the basic shapes which maximize covariation with the observed curves. In this way the probability of separating the effects of different sources of variability is maximized. The results of this step on the principal components in Figure 4 are presented in Figure 5.

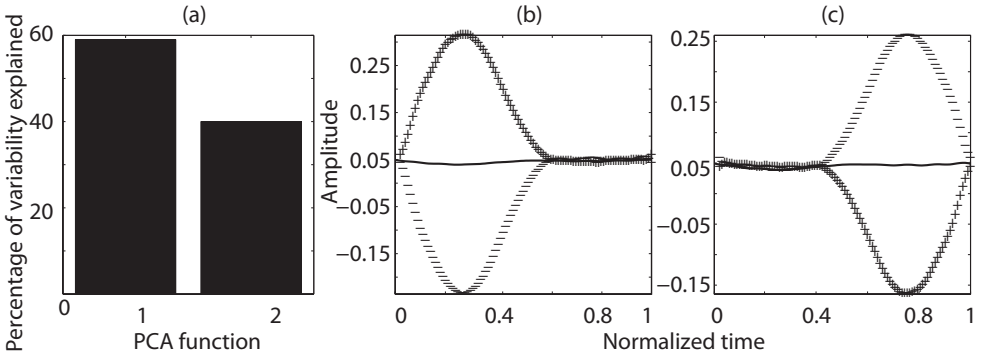

Figure 5: Panel a: Variability accounted by the first two rotated principal components. Panels b and c: effects of the first and of the second rotated principal components on the mean curve (see text).

Here each component introduces only one peak, as expected by adding one of the two curves from the left panel of Figure 3 to the average curve. Let's note that the mean curve is an approximately straight horizontal line centered over the value of 0.05, which is expected because the curves where generated starting from a uniform random distribution between 0 and 0.1.

Turning to real data, we applied functional PCA to the curves aligned in section 2.2.1 and 2.2.2. In order to isolate the components of variability in the shape of the trajectories, we subtracted from each curve its mean and divided the resulting curve for its range. Panel a from Figure 6 shows the amount of variance explained by each of the first three principal components. From panel b of the same figure we can observe that the effect of the first principal component, when multiplied by a positive coefficient and then added to the average curve, is that of reducing the overall amplitude of the jaw movement. However the amplitude of the oscillation which is synchronous with the lower lip movement (the one in the middle of each panel) is reduced only slightly while the amplitude of the oscillation synchronous with the lower lip (the one which is

represented at the edges of each panel) is reduced to one third of its amplitude in the average curve. The second principal component (panel c),

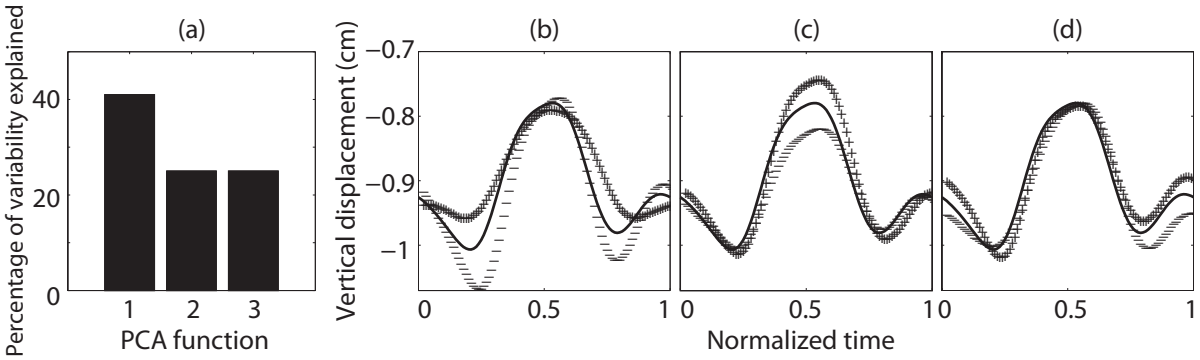

Figure 6: Effect of the Varimax rotated first three components on the mean trajectory of the jaw in the /pata/ utterances from Figure. 2

when multiplied by a positive coefficient, increases the amplitude of the jaw oscillation synchronous with the TTip movement. Finally, the third principal component (panel c), when multiplied by a positive coefficient, has a slightly more complex effect on the oscillation synchronous with the LLip movement. First, the peak height of this oscillation increases; second, the minimum which precedes the oscillation synchronous with the TTip is pushed down; third the minimum following the oscillation synchronous with the TTip is pulled up.

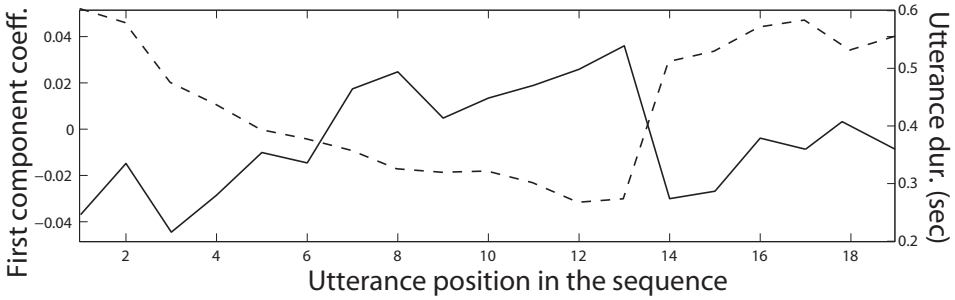

Figure 7: Evolution of the first principal component is plotted against the position of the utterance in a sequence. The continuous line indicates the values of loadings of the first principal component; the dashed line indicates the durations of the utterances measured in time steps. We can observe that as the duration of the utterances gradually decreases (up to the 13th repetition), the loading of the first principal component gradually increases. Between the 13th and the 14th repetitions there is an abrupt return to the initial values for both utterance duration and loading of the first principal component.

In Figure (7), we can see that the coefficient of first principal compo-

nent, interpreted as tracking the reduction from two to one jaw cycles per sequence, decreases in the middle portion of this sequence of /pata/ utterances (where the speech rate is higher). This is consistent with the reduction from two jaw cycles per utterance to one jaw cycle observed by Rochet-Capellan and Schwartz (2007) at fast speech rates. Moreover, at least in the analyzed sequence, the reduced jaw cycle is the one which is synchronous with the oscillation of the LLip.

3 Part II: Spectral methods

3.1 Introduction

The example outlined in the preceding section concerns the trajectory of the vertical position of the jaw observed in isolation, i.e. without considering its interactions with the other articulators. We are often interested in the relation between two or more variables, as when we study the coordination between the movements of two different articulators. However, if we are interested in their coordination, it is simpler and more appropriate to adopt spectral methods and study their instantaneous relative phase (Torrence and Compo, 1998).

At a given instant, the phase of a signal with cyclic (i.e. periodic) behavior is given by its position in its cycle. The phase is an angular variable and varies from 0, at the beginning of the cycle, to 2π at the end of the cycle. The relative phase between two cyclic signals is the difference between the positions of two oscillators in their relative cycles. If for example, the two oscillators start and end their own cycles at the same time, they are said to be in an in-phase relation (relative phase = 0); if the first oscillator is in the beginning of its cycle at the same time when the second one is in the middle portion of its own cycle the two oscillators are said to be in an antiphase relation (relative phase = π). If the phase angle lies between 0 and π, the first oscillator precedes the second one. When the phase angle lies between π and 2π, the first oscillator lags the second one.

In order to compute the instantaneous phase of a signal or the relative instantaneous phase between two signals, the original time-series must be represented in the time frequency domain. In other words, a spec-

tral representation must be adopted. When we adopt a spectral representation of a signal, we assume that the signal under study can be decomposed into different cyclic trajectories each one characterized by a different frequency of oscillation. The signal is thus approximated by a linear combination of analytic functions which oscillate with different frequencies. Since the characteristic evolution of the analytic functions is known, their instantaneous phase can be extracted. When looking for the temporal relation between two signals, we identify the time scale at which the two signals share more energy (their components at that frequency have higher amplitudes) and then the relative phase of the corresponding components can be computed.

3.2 Wavelet transform

The decomposition of a signal into several oscillators characterized by different frequencies is commonly achieved through the Fourier transform, which decomposes the signal into a linear combination of sine and cosine curves. To obtain the coefficients of the sinusoidal components in the Fourier transform, a section of the signal, whose length is fixed by the choice of an analysis window, is extracted. That section is compared with different combinations of sine and cosine functions which oscillate at different frequencies but have all the same length (i.e. the length of the section). The magnitude of the coefficients depends on the similarity between the sinusoids and the analyzed section of the signal. Once the coefficients of the sinusoidal components have been computed for a given section, the analysis window is shifted by a small amount (the time step of the analysis) and the computation of the coefficients is conducted for the next section. This approach has an important limitation in the fact that a single window is used to resolve all frequencies, implying that the temporal resolution of the analysis is the same in each point of the time-frequency plane. The consequence is that low frequencies produce too few oscillations into one window, affecting the localization of their frequency. At the same time, high frequencies present too many oscillations into a single window affecting the temporal localization of the signal's features.

The wavelet transform approach does not share this limitation because the analytic function (called the mother wavelet, cf. Figure 8) is scaled

(stretched or compressed in time) before being compared with the signal (however cf. section 5.2).

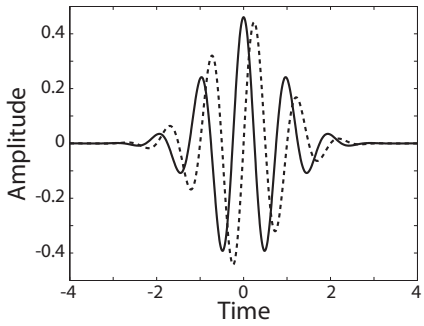

Figure 8: Shape of the complex Morlet wavelet (used as a mother wavelet for the analysis described in this chapter). Note that it is equivalent to a complex sinusoid modulated by a bell shaped curve to have positive energy on a continuous but limited interval of time.

Longer wavelets will be used in conjunction with low frequencies of oscillation, while shorter scales are used in conjunction with high frequencies of oscillation. In this way the wavelet transform can be sensitive to fast local changes in the signal, while maintaining a good separation of the energy at different frequencies.
As for the Fourier power spectrum, a spectral representation is obtained by squaring the absolute values of the coefficients (cf. Figure 9). The instantaneous phase values corresponding to the different frequency components are obtained from the complex argument of the corresponding coefficients. The wavelet equivalent of the Fourier power spectrum is called a scalogram. Time is represented on the x axis, while the y axis represents the scale factor used to stretch or compress the mother wavelet. However it is possible to characterize a compressed wavelet with the frequency of oscillation of the equivalent complex sinusoid (cf. Figure 8). A frequency (or period) scale on the y axis is thus often used to obtain more interpretable results. Figure 9a represents the power spectrum of the oscillation produced by the vertical position of the TTip sensor during a sequence of /pata/ utterances. Dark shades of gray are associated with strong energy. The concentration of the energy on the y axis indicates the oscillatory frequencies of the main components. The cone-shaped dashed line delimits the region in the time-scale plane where the wavelet transform gives reliable results (i.e. unbiased

by edge effects); this region is called the cone-of-influence (COI; Torrence and Compo, 1998).

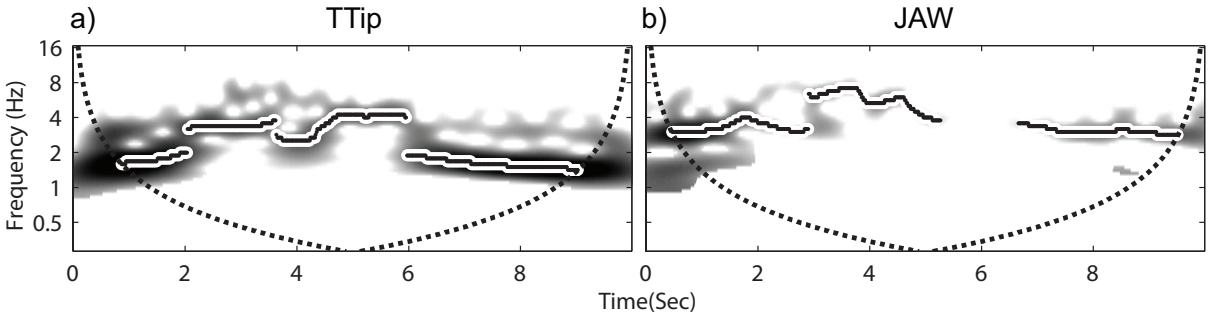

Figure 9: (a) Power spectrum of the oscillations produced by the vertical motion of the TTip during a sequence of repeated /pata/ utterances. (b) Power spectrum of the oscillations produced by the vertical motion of the jaw. The shades of gray represent spectral energy and correspond to the squared coefficients of the scaled Morlet wavelets. Time is represented on the abscissa. The frequencies of the sinusoids corresponding to the different wavelet scales are reported on the ordinate. Time and frequency resolution of the analysis vary over the ordinate. At high frequencies, high temporal resolution and low frequency resolution are observed. The inverse relation between time and frequency resolution holds at low frequencies. Dashed lines show COI; results exterior to this are influenced by edge effects.

Only energy values located within the COI are used in subsequent steps. The bold black line with white outline tracks the evolution over time of the strongest oscillatory component. At the beginning and at the end of the sequence (where the speech rate is low) we can observe a second strong oscillatory component, whose oscillations are twice as fast as the main component's oscillations. This is due to the coupling of the tongue tip with the jaw, whose spectrum is shown in Figure 9b. The duration of a cycle of the secondary component in the TTip spectrum is indeed equal to the duration of a jaw cycle.

The effects of the mutual influence between the jaw and the TTip are observed even in the spectrum for the jaw. Even here, in the regions corresponding to moderate speech rate, two components are distinguishable. Another important aspect of the jaw's spectrum is the weakening of the energy between 5.5 and 6.5 seconds. This event is signaled in the figure by an interruption of the line which tracks the energy of the main component. In the corresponding portion of the sequence, the spectrum energy is not significantly stronger than the spectral energy obtained by

a random process. This is consistent with the "freezing" of the jaw observed by Fuchs et al. (2009) at fast speech rates.

3.3 Cross wavelet transform

The product of the power spectrum of a signal with the complex conjugate of the power spectrum obtained from a second signal returns their cross spectrum, which represents the amount of energy shared by two signals at different frequencies of oscillation. The cross power spectra of the TTip and the LLip and of the jaw and the TTip are shown in panels a and b of Figure 10. At fast speech rates, only one component is present. At slow speech rates (t<2 sec. and t>6 sec.) two components are present. This is shown by the presence of two distinct horizontal gray bands in the regions of the plot. The stronger component corresponds to the lowest band (at a frequency slightly lower than 2Hz); the weaker component lies at a frequency of around 3Hz.

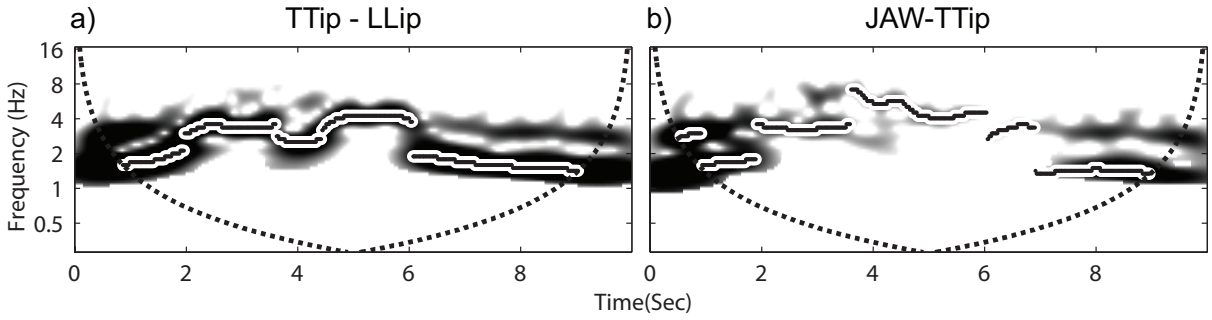

Figure 10: (a) Cross-spectrum of TTip and LLip oscillations; (b) Cross-spectrum of jaw and TTip oscillations

3.4 Obtaining relative phase relations

The relative phase between two signals can also be obtained from the cross spectrum. For each time point, we obtain a whole set of relative phase values characterizing the phase relations at different frequencies. A further step is thus needed to select, at each time point, the frequency of oscillation with the highest-valued coefficients (the band at which the two signals share the most of their energy), and then the corresponding phase value is chosen to characterize the instantaneous relative phase between the two signals at that point in time.

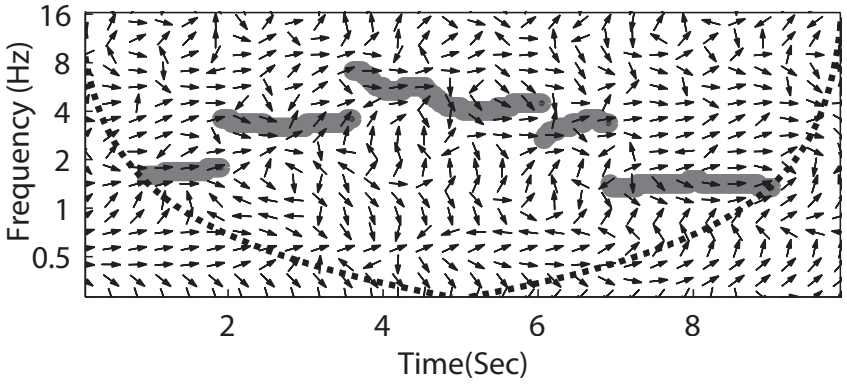

Figure 11: Phase plot from the spectrum in panel b of Figure 10

The phase plot of the spectrum in panel b of Figure 10 is shown in Figure 11, following Grinsted et al. (2004). The arrows indicate the relative phase at different frequencies (y axis) and different points in time (x axis). The phase angle between the TTip and the jaw is represented by the counter-clockwise angle between the arrows and the horizontal axis. The relevant phase values are those corresponding to the peaks of spectral energy which are indicated by the gray stripe on the background of the figure. It can be observed that the jaw is in phase with the TTip over all the duration of the sequence. This means that when two cycles per utterance are present, the strongest one will be in-phase with the TTip. When only one cycle per utterance is observed this is still in phase with the TTip.

3.5 Hilbert transform

Discrepancies between the shape of the observed signal and the shape of the mother wavelet may influence the outcome of the wavelet transform. This issue may lead to erroneous results in the presence of noisy signals. An alternative method, based on the Hilbert transform (Rosenblum and Kurths, 1998), should in principle reduce the effect of this kind of artifact and its use will be briefly described here (however cf. Bruns, 2004, for a demonstration of the substantial equivalence of the two approaches). The Hilbert transform, like the Fourier transform and the wavelet transform, is an analytical method; i.e., the signal is represented as an analytic function. The limitation of the Hilbert method is that it cannot handle multifrequency signals. For this reason the two time series are first sub-

mitted to a filter bank, obtaining two sets of band passed signals. The Hilbert transform is then applied to pairs of corresponding band passed signals. In this way both the time dependent relative phase and shared energy can be obtained for each frequency band. As with the wavelet transform the relevant values of the relative phase are those occurring at the shared energy peaks.

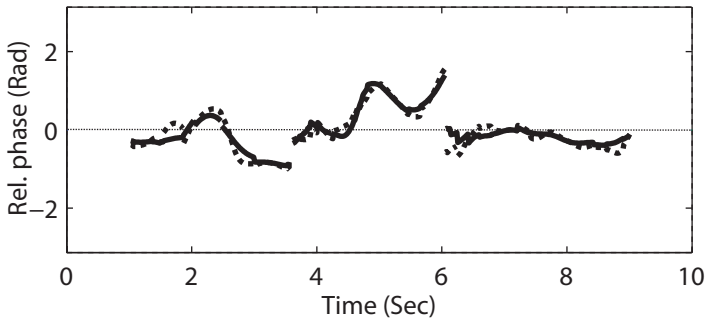

Figure 12: Relative phase between the trajectories of the jaw and the TTip used to produce the spectrum in Figure 10b. The continuous line represents phase values obtained using the cross wavelet transform; the dotted line represents values obtained using the Hilbert transform.

4 Part III: Recurrence and cross recurrence analysis

4.1 Introduction

Up to now we have explored methods to compare low dimensional signals or to study the relations between the components of a multidimensional system. What about comparisons among several high dimensional signals? We may ask for example how similar trajectories of jaw, TTip and LLip might be, across all components of their movement. Or we may want to evaluate the similarity of the evolution over time of the acoustic spectra corresponding to different utterances.
FDA is often used in the comparison of low dimensional time series which contain the same number of events (same number of peaks and valleys). However, the alignment procedure rapidly loses precision when the number of dimensions becomes high ($\gg 2$) and if the number of peaks and valleys varies between the time series. Recurrence analysis

methods (Maizel and Lenk, 1981; Eckmann et al., 1987; Webber and Zbilut, 1994; Marwan et al., 2007) offer an alternative approach that can deal effectively with high dimensional non-stationary signals. Such techniques have been developed within the framework of dynamical systems theory to study the stability of a potentially multidimensional signal which repeats its trajectory over time, but they have also been extended to the study of the behavior shared by different systems. In the former case, when the object of study is a signal which repeats its trajectory over time in a more or less regular fashion, recurrence analysis considers the distribution over time of the portions of the signal which are coherently repeated. In the comparisons of two signals, cross recurrence analysis focuses on the distribution over time of the subsections in which the signals show the same behavior (Zbilut et al., 1998; Marwan et al., 2002; Marwan and Kurths, 2002).

4.2 Method

We represent a multivariate time series as a time-varying sequence of state vectors, with each state vector containing the values observed on all the different dimensions of the time series. A recurrence plot (RP) is a two dimensional representation of those points in time when a state is repeated. Both axes of the plot represent the time scale of the signal and a black dot contained at position (i, j) means that the same state is observed at i and j positions in time (cf. Figure 13). The ith column of the plot contains the outcomes of the comparisons among the ith coordinate vector and all the other vectors of the time series. It follows that the plot is symmetric with respect to the main diagonal. This means that a dot at position (i, j) implies the presence of a dot at position (j, i).
On the main diagonal line we will always find black dots in a recurrence plot. This is due to the fact that on the main diagonal each state vector is compared to itself and the distance is thus 0. This line is therefore called the line of identity (LOI).
An isolated black dot at position (i, j) indicates that the state observed at position i is observed again at position j. However a diagonal line going from (i, j) to $(i + n, j + n)$ indicates that the state found at i is observed again at j, that the state found at $i + 1$ is observed again at $j + 1$, and so on, up to the state $i + n$. In other words, a continuous diagonal line, i.e. a

straight line of slope 1, indicates that a section of trajectory, whose length is equal to the length of the diagonal line divided by $\sqrt{2}$, is repeated in the signal. The bivariate signal represented in the figure is periodic and the same trajectory is perfectly repeated three times. In the RP, on each side of the LOI, we can thus observe two diagonal lines. [2]

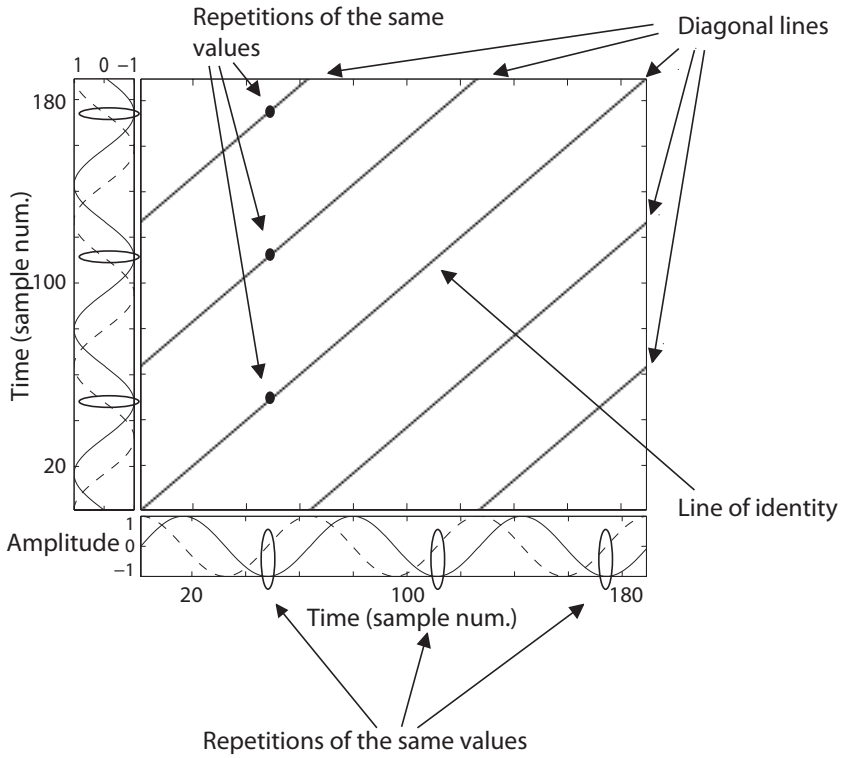

Figure 13: Top: recurrence plot of the bivariate time series plotted near the axes (see text).

2 Readers familiar with recurrence analysis may have noticed the absence of any reference to an embedding procedure. Such processing step is meant to increase the number of dimensions of the observed time series and it is motivated by the assumption that when observing a process we usually do not measure all its relevant features. Let's make the case that we are observing a process which is defined in three dimensions (for example, these may be the heights of the TTip, the LLip and the jaw in the production of our CVCV utterances) and that we can measure only the movement of one of the articulators involved (as if we were measuring the lip movement by means of a camera system such as Optotrack). Adopting an embedding strategy we can in principle reconstruct a three dimensional time series which shares several dynamical properties with the time series obtained by the recording of all the three articulators. The first dimension of the surrogate time series is represented by the recorded LLip height. The other two dimensions are copies of the first one but delayed in time. Although embed-

4.3 Cross recurrence plots

If the state vectors of a signal X are compared to the state vectors of a signal Y, we obtain what is called a cross recurrence plot (CRP; Marwan and Kurths, 2002). Each axis of the cross recurrence plot represents the time scale of one of the signals and the plot will show the points in time where the two signals exhibit the same state vector. A black dot at position (i, j) means that the same state is observed at position i in one signal and at position j in the other signal.

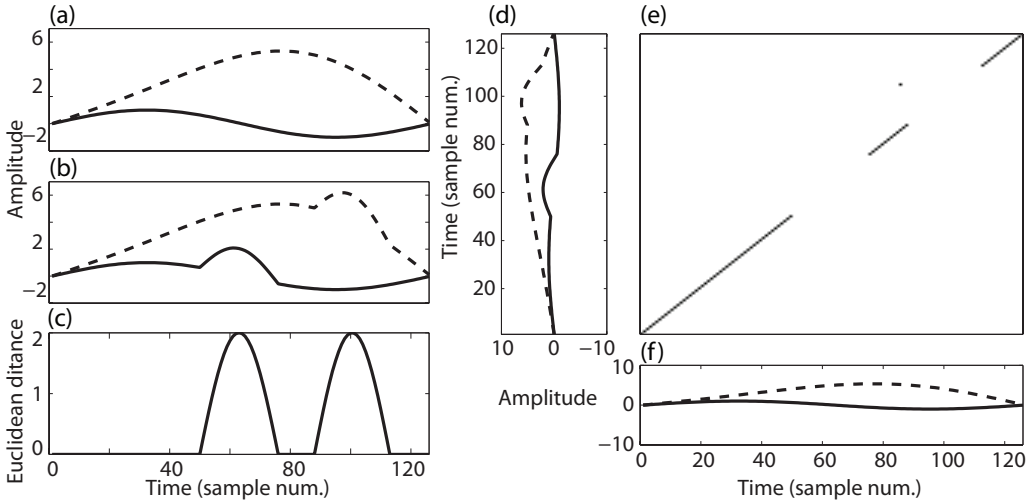

Figure 14: The time series shown in panels (a) and (b) and replicated in panels (d) and (f) are compared trough the cross recurrence plot in panel (e). The evolution of their Euclidean distance is plotted in panel (c).

The two bivariate time series presented in panels a and b of Figure 14 are compared through the CRP in panel e. Each trajectory is projected onto its axis in the CRP (panels d and f). The second trajectory has been obtained from the first one by slightly modifying its amplitude. The effect of the amplitude modification over the duration of the trajectory is

ding is essential for the analysis of mono-dimensional time series, it can be skipped when analyzing multi-dimensional time series. The lack of embedding potentially introduces artifacts in the recurrence plots (mainly the presence of lines with negative slope), however these artifacts are removed by the application of the algorithm presented in section 4.5. However to perform the embedding, two critical parameters have to be established: the number of dimensions to reconstruct and the delay. The computation of these parameters is not a trivial issue; moreover the whole procedure is sensitive to non-stationarity in the time series. Indeed for non-stationary signals an appropriate value for the delay parameter cannot be defined (Marwan, 2010). All this considered, we chose not to use embedding when comparing multivariate time series.

shown in panel c. Here the absolute difference between the time series is averaged over the two dimensions and it is plotted point by point. The two time series are maximally different between frames 50 and 75 and between frames 90 and 110.
A black dot at position (i, j) in the CRP means that a state vector at position i in the horizontal trajectory is replicated at position j in the vertical trajectory. The straight diagonal starting from the lower left corner of the plot indicates that the first 15 points of the horizontal time series match the first fifteen points of the vertical time series. Then a gap is encountered indicating the first mismatch between the time series. Notice that the other diagonal lines present in the CRP correspond to the region where the distance between the time series is smaller, and the gap is consequently located in the regions of maximal mismatch. If a black dot is isolated, i.e. surrounded by white regions, it indicates a random match between the two signals. In a cross recurrence plot, the main diagonal is not necessarily filled with black dots, because points at the same position in the two signals do not necessarily correspond.
Apart from isolated black dots and straight diagonal lines of slope 1, several other structures composed by connected dark dots can be found. Importantly the slope of the continuous lines varies as a function of the relation between the time scales of the two signals. Indeed these slopes correspond to the ratio between the two rates of change of the matching parts of trajectories. A slope which is smaller than one indicates that the signal whose time scale is on the vertical axis is faster, a slope which is greater than one indicates that the signal whose time scale is represented on the horizontal axis is faster. Bowed continuous lines indicate that the relation between the two time scales changes over time. If a bowed quasi-continuous line is found around the main diagonal, this is called the line of synchrony (LOS). Indeed to synchronize the two signals, it suffices to shift the position in time of the points of one signal by an amount which is equal to the vertical distance between the corresponding points on the LOS and the main diagonal of the plot (i.e. the line connecting the opposite corners).

4.4 How a cross recurrence plot is built

Each of the two signals X and Y is represented by N state vectors and each state vector represents the M coordinates of the system at a given point in time. A distance matrix is computed by comparing each state vector of the signal X to each state vector of the signal Y. The vertical side of that matrix has the same length as the signal Y, while the horizontal side is as long as the signal X. Each cell of the matrix contains a measure of the difference between the state vector of the first signal and the state vector of the second signal. If the Euclidean distance is adopted, this will be equal to the square root of the summed squares of the distances computed on the different dimensions:

$$\left\| X_i - Y_j \right\| = \sqrt{\sum_{k=1}^{m} \left(X_{i,k} - Y_{j,k} \right)^2} \tag{5}$$

where m is the number of dimensions of the two signals X and Y. In the construction of a CRP a maximum distance is often used as it is generally faster to compute; i.e. the distance between two state vectors corresponding to the maximum among the absolute values of the distances computed on the different dimensions.

$$\left\| X_i - Y_j \right\|_\infty = argmax \left(\sqrt{(X_{i,1} - Y_{j,1})^2}, \cdots, \sqrt{\left(X_{i,m} - Y_{j,m} \right)^2} \right) \tag{6}$$

The differences are stored in the matrix in the following way: the first column holds the differences between the first state vector of the signal X and all the state vectors of the signal Y, going from the bottom to the topmost row, where the first state vector of X is compared to the last state vector of Y. The distances from the other points of the signal X are stored in a similar fashion in the columns which follow the first one. The final plot is then derived by the distance matrix, deciding, for each distance measure, if it is sufficiently small such that the compared state vectors can be considered equal. With a fixed threshold criterion, each point of the plot corresponding to a distance smaller than a predetermined threshold will be represented by a dark dot. All the other points will be represented by white dots. The selection of an appropriate threshold is aimed at finding an equilibrium between having a threshold which is as

small as possible, while being capable to detect a sufficient number of recurrences. Thiel et al. (2002) show that the threshold should be 5 times larger than the standard deviation of the observational noise present in the time series. However, how to obtain the appropriate threshold in the general case where no estimate of the observational noise is available, remains an open issue (cf. Marwan, 2010, for a review of the relevant approaches).
In computing a cross recurrence plot different approaches involving an adaptive threshold can be adopted. For example, with the **fixed recurrence rate** criterion, the number of recurrence points to be found in the plot is fixed at a given percentage of the number of locations in the plot (corresponding to the maximum possible number of recurrence points given the lengths of the time series compared). The threshold is then adapted in order to obtain the desired number of recurrences. With a **fixed amount of neighbors** criterion, the threshold changes from one column of the plot to the other in such a way that the same number of recurrences is found in each column. In this way, each point of the time series represented on the horizontal axis of the plot has the same number of recurrences in the other time series.

4.5 Recurrence quantification analysis (RQA)

Quantitative indices of the relations between two signals are often derived by counts of the points belonging to diagonal lines of slope 1 (Marwan and Kurths, 2002). For example the percentage of determinism (%DET) is equal to the number of points belonging to diagonal lines divided by the total number of dark points present in the plot. This means that it is equal to the ratio between the number of points belonging to repeated trajectories and the number of the random matches. The mean length of the diagonals and their maximum length are used as well.
There are at least two reasons not to use these measures for signals derived from speech when comparing different realizations of the same motion pattern. One is related to the differences in their time scales. Indeed bowed lines are not captured by the measures described above. For this reason, the above measures are not able to separate variability in the amplitude of the time series from variability in their rate of change. This is particularly problematic if we want to compare different produc-

tions of the same utterance to measure variability (cf. Lucero, 2005). The second reason is related to the smoothness of the signals. If X and Y present a match at position (i, j), and X is particularly smooth, there are good probabilities that the two signals will also match at position $(i + 1, j)$ because of the small difference between $X(i)$ and $X(i + 1)$. A random match can easily be transformed in a continuous horizontal line, if the signal whose timescale is represented on the X axis is particularly smooth. For the same reason a continuous diagonal or bowed line can become thicker and thicker as the smoothness of the signal Y increases. While differences in the time scales of the two signals can obscure the similarity between sections of the two signals, smoothness introduces false matches. To reduce the bias introduced by these two features, in our recent work (Lancia, Fuchs and Tiede, in preparation) we introduce an algorithm based on image processing techniques combined with a skeletonizing strategy proposed by Marwan et al. (2002).
The main processing steps are the following:
1) Each group of connected dark dots is distinguished through a standard algorithm for connected components labeling (Ballard and Brown, 1982, pp. 13-62).
2) Each group is then reduced to a thick line which follows its shape. The tracking is constrained to proceed towards the top-right direction. This is achieved by first identifying the dark dot nearest to the bottom left corner of the smallest rectangle including the group. Then the tracking algorithm described by Marwan and his colleagues is used. In summary this algorithm starts looking for recurrence points by placing a square window with its bottom left corner on the starting point. The size of the window is increased until it includes at least another black dot. When this happens, the center of mass of the area under the window is computed and it is considered as the second point of the tracked line. A new window is placed with the bottom left corner onto the found point and its size is increased until a new dark point is found. The procedure is repeated until the tracked line reaches the right or the top side of the plot. The percentage of determinism is then computed by dividing the number of dark dots belonging to the thick continuous lines by the total number of possible dark dots.

4.6 Application to speech

We consider here 21 sequences of /pata/,/fata/ and /pasa/ utterances produced by three speakers[3]. In each sequence, we compare pairs of successive utterances using recurrence analysis and consider the %DET as a measure of similarity between the utterances of each pair. This choice was motivated by the assumption that two utterances produced one after the other are more likely to be similar than two utterances far away in the sequence. Comparisons among successive utterances are thus expected to better differentiate among levels of variability observed in different conditions.

Since we study the joint motion of the jaw, the TTip and the LLip, a six dimensional coordinate vector is needed to store the vertical and horizontal positions of the three articulators at different moments in time. Before performing the comparison each trajectory is down-sampled to a standard length. This step is performed in order to compare measures derived from different pairs of utterances[4].

The recurrence plot shown in Figure 15b is obtained by comparing two consecutive utterances. The actual multi-variate trajectories are plotted along their respective axes (panels a and d). To avoid crowding of trajectories, only the vertical positions of the articulators are displayed. In panel c we can observe the output of our algorithm when that recurrence plot is processed.

Note the presence of continuous lines going downward in the unprocessed RP. This means that a portion of the signal in Figure 15b is pulled in the opposite direction by the signal in 15a. This is reasonable, considering that the motion of the articulators has an oscillatory nature and that a movement in one direction is always followed by a movement in the opposite direction. However the points belonging to these lines do

3 In order to match the results from recurrence analysis with the results from functional PCA, we selected the three speakers in the following way: we extracted the first principal component separately for each speaker and utterance; we then chose speakers for whom the first principal component reduces the amplitude of the jaw cycle synchronous with the labial constriction. The loadings obtained in this way were used to compute the correlation shown in Figure 16.

4 Using the fixed recurrence rate criterion, the probability that the outcome of a comparison between two coordinate vectors produces a recurrence point depends on the number of coordinate vectors compared, i.e. on the lengths of the two time series.

not represent deterministic matches, in the sense defined in the preceding section, and thus they are correctly discarded by our algorithm.

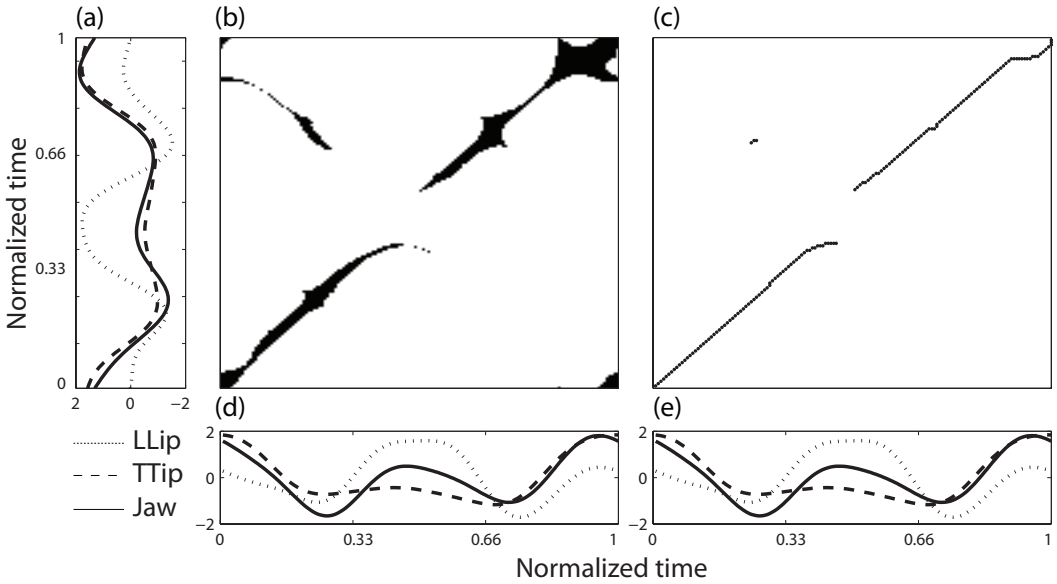

Figure 15: Cross recurrence plots of the 6 trajectories produced by the jaw, the TTip and the LLip in the horizontal and vertical dimensions during two successive utterances. Trajectories in the horizontal dimension are omitted for clarity. The first trajectory is plotted on the horizontal axis of each plot, the second trajectory is plotted on the vertical axis common to the two plots. The plot in panel b is the classic CRP, the plot in panel c is obtained by submitting the plot in panel b to the algorithm described in the text.

Once the comparisons have been performed and the %DET has been computed for all pairs of successive utterances, we obtain a sequence of values for each sequence of utterances. We will compare the values of the %DET index with the values of the coefficient relative to the first principal component of the jaw motion described in the first section. This comparison is motivated by the following reasoning:
We adopted the hypothesis that in this particular task, the labial coronal ordering is favored by the weakening of the jaw cycle which is synchronous to the movement of the LLip. Using functional PCA we observed that jaw behavior evolves from two jaw cycles per utterance to one cycle per utterance. This change is captured by the variation of the first principal component of the jaw's motion which, when averaged over the nine sequences, explains just under 90% of the total spatio-temporal variation. We could then observe that, at the beginning of the sequences, each utterance presents two jaw cycles of comparable amplitude and, after a transient portion during which speech rate increases,

the second behavior is observed, with a reduction of one of the two jaw cycles. If the behavior of the whole system changes because the initial behavior becomes unstable at high speech rates, we expect to observe that, for short utterance durations, %DET increases as the coefficient of the first principal component decreases. In fact, this is supported by the weak but highly significant negative correlation (r^2=-0.37, p<10^{-5}) shown in Figure 16 where the two indices are plotted against one another.

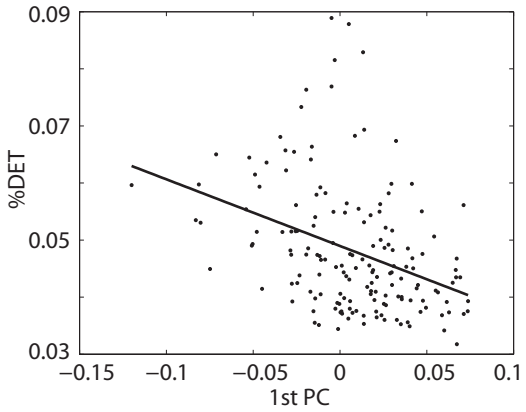

Figure 16: %DET regressed with the value of the coefficient of the first principal component obtained by applying functional PCA to 24 CVCV sequences obtained from three speakers r^2=-0.37, p<10^{-5}).

5 Part IV: General discussion and conclusions

5.1 Parameterization of the analysis

Functional data analysis: Given a set of curves and a shape for the basis function, a certain number of parameters are responsible for the quality of the fit of the smoothed curves to the original trajectories. The first parameter to be set is *the number of the basis functions* used to model the original observations. For the Fourier basis, this corresponds to the number of sine and cosine components plus one; for the B-spline basis, this corresponds to the number of segments in which the original curves are decomposed, plus the order of the polynomials used to model the different segments. In both cases, a higher number of basis functions determines an increase in the sensitivity to local features of the trajectories. An automatic method has been proposed to determine the optimal

number of basis functions needed to model a trajectory (Friedman and Silverman, 1989). However this approach does not always give useful results and often visual inspection remains the only way to evaluate the fit to the original data.
When using a B-spline basis, *the order of the polynomials* may be changed too. This affects the smoothness of the modeled curve and of its successive derivatives. The basis functions are constrained to match at the break points, and their successive derivatives up to the order minus two. Increasing the order of the polynomials increases the number of successive derivatives which are constrained to match the break points, and as a consequence the modeled curves appear smoother. When working with tangential velocity, the original observations should be modeled by means of splines of the fourth order or higher to obtain appropriately smooth derivatives.
The last parameter for the basis expansion represents *the emphasis given to the roughness penalty constraint* in the optimization algorithm which determines the weights of the basis functions. The higher the values for this parameter (λ; Ramsay and Silverman, 1997), the smoother the modeled curves will be. The optimal value for this smoothing parameter can be obtained by using a generalized cross validation criterion (Craven and Wahba, 1978). However, even for this parameter, it is necessary to inspect the results obtained with different values (Ramsay and Silverman, 1997).
When the functional observations need to be aligned, the same set of parameters must be controlled in the basis expansion of the warping functions. These curves are generally built starting from a B-spline basis of the fourth order. The general advice is to start with a very small number of bases (i.e. a small number of break points) and then to increase this number if the results of the registration are not satisfactory.
Cross wavelet transform: The wavelet transform is subject to the usual tradeoff between time and frequency resolution. In the implementation of the cross wavelet transform discussed in this chapter, the equation of the Morlet mother wavelet is defined by only one parameter which represents a dimensionless frequency. A good balance between localization in time and frequency is obtained using a value of 6 for this parameter (Torrence and Compo, 1998).
Recurrence analysis: In this chapter we used recurrence analysis with

a fixed recurrence rate criterion to determine if two state vectors of two different multivariate signals can be considered similar. With this choice, the parameter that controls the behavior of the analysis is *the ratio of dark dots to the number of possible dots* in the plot. If this value is too small, recurrence points can be missed; however if this value is too high, false recurrences may be detected. The choice of this parameter for a nonstationary signal is an unresolved question. Schinkel et al. (2008) propose to use the value which best separates the classes of trajectories to be studied. For example, in the case discussed in this chapter, we could have maximized the difference in the %DET index between curves which show high values for the first principal component and curves which show low values for this component. This maximization can be conducted computing the ROC curve from the obtained %DET index values. The algorithm outlined in this chapter for post-processing the recurrence plots reduces the impact of this parameter on the final results in that false recurrences are detected and discarded. When using such a method, is thus advisable to choose a value smaller than the 10% of the total number of possible dots in the plot.

5.2 Limits of this survey

The analyses described in this chapter cover only a small portion of the possibilities offered by the methods proposed. This is especially true for FDA and RQA.
Indeed several methods related to multivariate statistics and linear regression have been adapted to the analysis of functional data. Moreover FDA has been shown to be successful in the assessment of the dynamical laws underlying the evolution of time series (cf. Ramsay and Silverman, 2002, for a review of several variants and applications of FDA).
Concerning the spectral methods in section 3, it must be noted that, although the different approaches are formulated in different terms, all correspond to a convolution of a known function with the analyzed signal and can be implemented to give the same results (Bruns, 2004; Le Van Quyen et al., 2001; Quiroga et al., 2002; Kiebel et al., 2005). Even the Fourier transform can reach the same flexibility of the other methods when the window size is varied along the signal according to its local frequency. We chose to focus on the wavelet approach due to its ease of

application: no filter bank has to be implemented and no tracking of the frequency of the main component has to be conducted before applying the transform. The only preprocessing needed is a normalization of the signals compared.
Cross recurrence analysis has been used to detect different kinds of synchronization between time series (Marwan et al., 2007). Recurrence analysis has been used to recover the dynamical invariants of the processes underlying the time series and to detect the presence of chaotic behavior (Romano et al., 2005). However, these uses of recurrence analysis are controversial if the evolution over time of the system is not drawn in its reconstructed phase space, i.e. the space defined by the actual degrees of freedom of the system underlying the observed time series (Thiel et al., 2004).

5.3 General conclusions

In contrast with these advanced uses, the examples illustrated in this chapter are meant to represent situations in which the application of the techniques is quite straightforward and requires the smallest possible number of choices. FDA is a good alternative when the trajectories observed are low dimensional, smooth, and characterized by moderate variability (i.e. all the trajectories present the same number of peaks). Spectral methods as the cross-wavelet transform are useful for characterizing the coordination among different signals. Finally, recurrence analysis becomes particularly useful when comparing multivariate trajectories having high levels of variability.

6 Acknowledgments

This paper was made possible by the public availability of the software reported in Appendix B, made freely available through the efforts of several generous researchers. We wish to thank Pascal Perrier for much useful discussion, and Susanne Fuchs both for promoting the production of this chapter and for her contributions in the development and application of the method summarized in section 4.6. Finally, we would like to thank two anonymous reviewers for useful and important sug-

gestions. This work was supported by a grant from the German-French University to the PILIOS project.

7 Appendices

A Further readings and tutorials

Ramsay and Silverman (2002) offer a collection of case studies analyzed through FDA with a reduced emphasis on mathematical formulations. The commented Matlab and R implementations of the analysis described in this book (and the relative datasets) are available at this address `http://www.stats.ox.ac.uk/~silverma/fdacasebook/`.
Ramsay et al. (2009) wrote an introduction to FDA and its variants with examples in both R and Matlab languages. Ramsay and Silverman (1997) is the principal reference on FDA with a detailed discussion of the mathematical and computational aspects. An introductory tutorial with emphasis on the concept of spline smoothing is available at this address: `http://www.psych.mcgill.ca/misc/fda/examples.html`.
Lucero's first applications of FDA to study the variability of speech signals are published in Lucero et al. (1997). A tutorial-like presentation can be found in Lucero (2005). Gubian maintains a web page on functional data analysis for speech research (`http://lands.let.ru.nl/FDA/`). The available tutorials cover FDA registration, functional PCA and functional linear models.
A general tutorial on the use of cross wavelet transform has been written by Torrence and Compo (1998). Grinsted et al. (2004) build on the work by Torrence and Compo in their very accessible tutorial on the use of cross wavelet transform for the analysis of geophysical data.
Marwan (2003) provides a comprehensive introduction to recurrence and cross recurrence analysis, while Marwan et al. (2007) can be considered the most complete reference on recurrence analysis and its variants. Van Lieshout and Namasivayam (2010) report results obtained by using recurrence analysis with speech articulators[5]. However in that applica-

5 To be more precise, recurrence analysis was not conducted directly on the movements of the articulators. First the relative phase was computed from pairs of kinematic signal, then the recurrence analysis was applied to the time series of the relative

tion no method to compensate for non-stationarity was adopted. Non-stationarity implies that the rate of change of the signals analyzed varies, introducing bowed lines in the plot and, as a consequence of that, the results obtained in that paper are hard to interpret. Webber Jr. and Zbilut (2005) and Shockley (2005) wrote accessible tutorials on the use of recurrence and cross recurrence analysis. A web based tutorial can be found at the following address: `http://www.recurrence-plot.tk/glance.php`, in which an animated introduction presents in an intuitive way the concept of phase space reconstruction. This topic has not been covered in the present chapter but constitutes the basis for more advanced uses of this technique.

B Software

With the exception of our algorithm designed for post-processing cross recurrence plots, all the analyses described in this chapter can be performed through available Matlab and R commands. The FDA software page is maintained by Ramsay at this address: `http://www.psych.mcgill.ca/misc/fda/downloads/FDAfuns/`. Here Matlab, R, and Splus versions of the software can be found.
Torrence and Compo's cross wavelet transform software is available at: `http://paos.colorado.edu/research/wavelets/software.html`.
Grinsted et al. (2004) wrote a Matlab package that integrates Torrence and Compo's toolbox with several useful routines. One of these routines has been used to produce the phase plots presented in this chapter. The package can be found at this address: `http://www.pol.ac.uk/home/research/waveletcoherence/`. Marwan's Matlab toolbox for recurrence and cross recurrence analysis can be found at the following address: `http://tocsy.agnld.uni-potsdam.de/crp.php`.
Webber produced a series of programs to perform recurrence analysis in DOS/Windows environment. This can be found at the following address: `http://homepages.luc.edu/~cwebber/`.

phase values.

References

Aston, J. A. D., Chiou, J. M., and Evans, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):297–317.

Ballard, D. H. and Brown, C. M. (1982). *Computer Vision*. Prentice-Hall, New York.

Bruns, A. (2004). Fourier-, hilbert-and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, 137(2):321–332.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403.

Eckmann, J. P., Kamphorst, S. O., and Ruelle, D. (1987). Recurrence plots of dynamical systems. *EPL (Europhysics Letters)*, 4:973–977.

Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics*, 31(1):3–21.

Fuchs, S., Hartinger, M., Ziervogel, M., and Weirich, M. (2009). On coordinative patterns and transformation effects in reiterant speech. In *Proceedings of the Conference Progress in Motor Control, Marseille*.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3):386–412.

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11(5-6):561–566.

Gubian, M., Cangemi, F., and Boves, L. (2010). Automatic and data driven pitch contour manipulation with functional data analysis. In *Proceedings of Fifth International Conference on Speech Prosody, Chicago, USA*.

Hertrich, I. and Ackermann, H. (2000). Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *The Journal of the Acoustical Society of America*, 107:2236–2247.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.

Keating, P. A., Lindblom, B., Lubker, J., and Kreiman, J. (1994). Variability in jaw height for segments in english and swedish vcvs. *Journal of Phonetics*, 22:407–422.

Kiebel, S. J., Tallon-Baudry, C., and Friston, K. J. (2005). Parametric analysis of oscillatory activity as measured with eeg/meg. *Human Brain Mapping*, 26(3):170–177.

Koenig, L. L., Lucero, J., and Löfqvist, A. (2003). Studying articulatory variability using functional data analysis. In *Proceedings of the 15 th International Congress of Phonetic Sciences*, pages 269–272.

Lancia, L. and Fuchs, S. (2011). The labial coronal effect revisited. In *Proceedings of the ninth International Seminar on Speech Production, Montreal*.

Le Van Quyen, M., Foucher, J., Lachaux, J. P., Rodriguez, E., Lutz, A., Martinerie, J., and Varela, F. (2001). Comparison of hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of Neuroscience Methods*, 111(2):83–98.

Lindblom, B. (1983). Economy of speech gestures. In MacNeilage, P., editor, *The Production of Speech*, pages 217–245. Springer, New York.

Lucero, J. C. (2005). Comparison of measures of variability of speech movement trajectories using synthetic records. *Journal of Speech, Language, and Hearing Research*, 48(2):336—-344.

Lucero, J. C. and Koenig, L. L. (2000). Time normalization of voice signals using functional data analysis. *The Journal of the Acoustical Society of America*, 108:1408–1420.

Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*, 40(5):1111–1117.

Maizel, J. V. and Lenk, R. P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the National Academy of Sciences*, 78(12):7665.

Marwan, N. (2003). *Encounters with Neighbours - Current Developments of Concepts Based on Recurrence Plots and their Applications*. University of Potsdam.

Marwan, N. (2010). How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos*, 21(4):1003–1017.

Marwan, N., Carmen Romano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329.

Marwan, N. and Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, 302(5-6):299–307.

Marwan, N., Thiel, M., and Nowaczyk, N. R. (2002). Cross recurrence plot based synchronization of time series. *Nonlinear Processes in Geophysics*, 9(3-4):325–331.

Mooshammer, C., Hoole, P., and Geumann, A. (2007). Jaw and order. *Language and Speech*, 50(2):145–176.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the slip technique. *Language and Speech*, 50(3):311–341.

Quiroga, R., Kraskov, A., Kreuz, T., and Grassberger, P. (2002). Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Physical Review E*, 65(4):041903.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (1997). Functional Data Analysis.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Rochet-Capellan, A. and Schwartz, J. L. (2007). An articulatory basis for the labial-to-coronal effect:/pata/seems a more stable articulatory pattern than/tapa. *The Journal of the Acoustical Society of America*, 121:3740–3754.

Romano, M., Thiel, M., Kurths, J., Kiss, I., and Hudson, J. (2005). Detection of synchronization for non-phase-coherent and non-stationary data. *EPL (Europhysics Letters)*, 71(3):466–472.

Rosenblum, M. and Kurths, J. (1998). Analysing synchronization phenomena from bivariate data by means of the hilbert transform. In Kantz, H., Kurths, J., and Mayer-Kress, G., editors, *Nonlinear Analysis of Physiological Data*, pages 91–99. Springer, Berlin.

Schinkel, S., Dimigen, O., and Marwan, N. (2008). Selection of recurrence threshold for signal detection. *The European Physical Journal-Special Topics*, 164(1):45–53.

Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. In Riley, M. A. and Van Orden, G. C., editors, *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences, Web Book*, pages 142–177. National Science Foundation, Arlington, Virginia.

Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., and McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research*, 104(3):493–501.

Thiel, M., Romano, M. C., Kurths, J., Meucci, R., Allaria, E., and Arecchi, F. T. (2002). Influence of observational noise on the recurrence quantification analysis. *Physica D: Nonlinear Phenomena*, 171(3):138–152.

Thiel, M., Romano, M. C., Read, P. L., and Kurths, J. (2004). Estimation of dynamical invariants without embedding by recurrence plots. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 14(2):234—243.

Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Van Lieshout, P. and Namasivayam, A. K. (2010). Speech motor variability in people who stutter. In Maassen, B. and van Lieshout, P., editors, *Speech Motor Control: New Developments in Basic and Applied Research*, pages 190–214. Oxford University Press.

Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3):1251–1276.

Webber, C. L. and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76(2):965–973.

Webber Jr., C. L. and Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In A., R. M. and C., V. O. G., editors, *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences, Web Book*, pages 26–94. National Science Foundation, Arlington, Virginia.

Zbilut, J. P., Giuliani, A., and Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, 246(1-2):122–128.