Functional PCA for F_0 contour analysis

Michele Gubian

May 25th, 2014

This document provides an overview of the method proposed in Gubian et al. (2011) for the analysis of F_0 contours. The method is based on Functional Principal Component Analysis (FPCA), which is one of the data analysis algorithm available within the Functional Data Analysis (FDA) framework. This document consists of four sections, the first three of them corresponding to the three main steps of the procedure, namely smoothing, landmark registration and Functional PCA, respectively. The last section points out that the method illustrated here is not the only one available in the literature, and some references to other approaches and related software packages are provided. Readers interested in applying FDA to their own data are invited to visit the website maintained by the author[1], where didactic material and code recipes can be downloaded. A comprehensive account on FDA can be found in Ramsay and Silverman (2005); Ramsay et al. (2009).

1 Smoothing

Smoothing transforms a contour composed of discrete samples into a continuous curve represented by a mathematical function (this is what "functional" refers to in FDA). The target function is chosen from a set of possible functions specified by the user. In the case of F_0 contours, which can assume a wide range of shapes, it is customary to adopt *B-splines* as the general function set (de Boor, 2001). A B-spline is a sequence of polynomial curves that summed together approximate the desired contour. By varying the number of polynomial components of B-splines and a parameter influencing the curve smoothness, the user controls how close to the original discrete samples the

[1] `http://lands.let.ru.nl/FDA`

continuous curve will be, or in other words, the user controls the temporal resolution of the continuous representation (Ramsay and Silverman, 2005). An example is shown in Figure 1, where the smoothing curve was chosen in such a way to eliminate some microprosodic detail present in the first part of the original contour. In general it is important to find a good compromise between smoothing too much, which would delete relevant information from the original contours, and not smoothing enough, which would leave irrelevant detail that makes the extraction of global trends harder in the following steps.

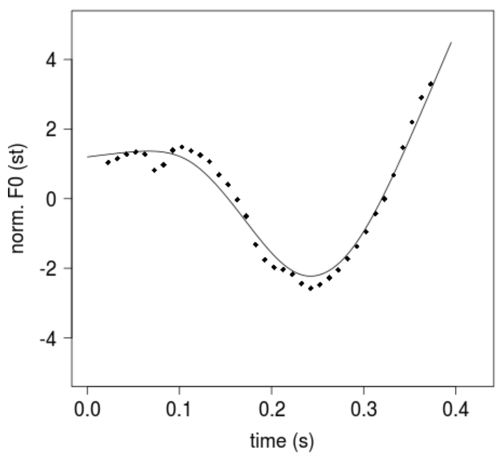

Figure 1: Example of a smoothed F_0 contour. Dots represent F_0 samples obtained from the pitch tracker available in Praat. The curve is a B-spline. This contour is extracted from a realisation of a three-syllabic word. The y-axis reports F_0 values in semitones after the global mean value was subtracted (thus corresponding to the zero level).

The main conceptual aspect concerning smoothing is that it provides an entry point for prior knowledge. In our case, selective inclusion or exclusion of prior hypotheses on the nature of the input F_0 contours is realised in two ways. First, the choice of B-splines implies that no prior hypothesis is introduced in terms of the expected global shape of the contours, e.g. the number and position of rise-fall movements is not fixed in advance, and can vary across contours. Second, the user can always determine the degree of smoothing, which can reflect prior knowledge on the time resolution of interest. In our case, Figure 1 shows that we opted for a curve that is smoother than the one we would need in order to model the input samples exactly, i.e. to produce a curve that intercepts all samples. This is because we are not interested in reproducing details that have to do either with microprosodic effects or possibly with measurement errors of the F_0 tracker.

2 Landmark registration

The analysis of intonational events, like pitch accents, is generally based on descriptions of the F_0 movement in relation to the underlying segments, e.g. syllables or morae. Typically, a data set collected to study a specific intonational phenomenon is made of F_0 contours extracted from realisations of the same utterance, or of utterances that share the same segmental structure. In this way, systematic differences in the shape of contours can be attributed to the (linguistic) factors relevant for the study, since the segmental material does not change (e.g. we want to be able to describe the behavior of F_0 in relation to the onset of the nuclear syllable of a word in different linguistic conditions).

FDA provides a way to alter the shape of curves in order to align a set of events (*landmarks*) defined by the user that correspond to each other on every curve. In the analysis of F_0 contours, this facility is used in order to synchronise all corresponding segmental boundaries. The purpose of this operation, called *landmark registration*, is to be able to interpret the results of FDA in terms of the segmental structure underlying the contours. If registration was not done, at the end of the analysis we would not be able to distinguish contour shape variations that are a consequence of a variation of segment durations (e.g. a peak occurs always inside the same segment, but the preceding segments vary their duration) from variations of pitch alignment (e.g. a peak occurs before or after a given segmental boundary). Landmark registration alters the time axis of a curve in such a way that landmarks move from their original position to a new position specified by the user; usually the average position computed on the whole curve set is chosen as common location of landmarks. The operation is carried out automatically and it is only based on the time position of the selected landmarks, where the latter can obtained either from the manual annotation e.g. carried out in Praat, or automatically using a speech recognition software (e.g. see Turco and Gubian, 2012). The time warping involved in landmark registration is guaranteed to preserve the qualitative aspects of curves, i.e. all rise and fall movements are preserved in the original order and level, no jumps or ruptures are produced. The left panel in Figure 2 shows an example.

The output of landmark registration is a new set of curves where all corresponding segments have the same duration, which means that the information on the differences in segment durations is discarded. Gubian et al. (2011) proposed a way to represent the original durations my means of functions (curves) and to incorporate this representation in the following statistical analysis. Let $h(t)$ be the time warping curve produced by landmark registration that specifies the mapping between normalized time t and original time $h(t)$ for a specific normalized curve $F_0(t)$ (bottom left panel in Figure 2). This curve contains all the information necessary to reconstruct the original $F_0(t)$ contour (solid curve in top panel of Figure 2) from its registered version (dashed curve ibid.). An equivalent representation of $h(t)$ that is more convenient for the statistical analysis tools that follow is obtained by

$$r(t) = -\log \frac{dh(t)}{dt}. \tag{1}$$

The curve $r(t)$ represent the log of the relative rate of realisation of the utterance with respect to its normalised version. According to Eq. (1), if a segment (i.e. anything between two consecutive landmarks) is realised two times faster than in the normalised version, then $r(t)$ will read values around $\log 2 = 0.7$ along the time interval corresponding to that segment; vice versa, values around $\log \frac{1}{2} = -0.7$ will be reached for segments pronounced at half rate (i.e. double duration). The bottom right panel in Figure 2 shows curve $r(t)$ obtained from $h(t)$ in the bottom left panel. The advantage of using $r(t)$ instead of $h(t)$ is that the former represents proportional variations in duration linearly. Fortunately, since Eq. (1) is completely reversible, once we obtain results in terms of $r(t)$ we can convert them to $h(t)$ values, which correspond to landmark positions, or directly to differences between $h(t)$ values, which correspond to segment durations. This will be shown in the next section.

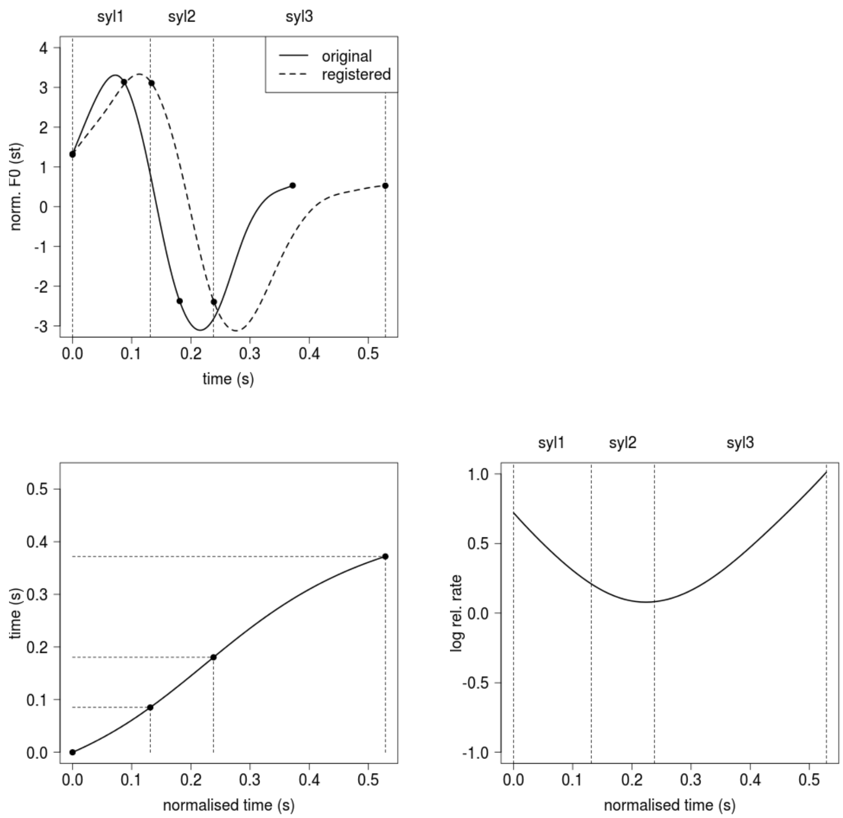

Figure 2: Top: an F_0 curve extracted from a realisation of a three-syllabic word (solid curve) and its landmark registered version (dashed curve). Dots on the curves mark the position of the landmarks (here syllabic boundaries), and vertical dashed lines show the position of the registered landmarks, which correspond to the average position computed on all the collected realisations of this word (93 in this case). Bottom left: the time warping curve $h(t)$, which expresses the relation between the original and the registered curve in the left panel. Bottom right: the log relative rate curve $r(t)$, which is obtained from $h(t)$ by applying Eq. (1). In this case, the segments in the original utterance are shorter than in the registered version, thus the values of $r(t)$ tend to be above zero.

3 Functional Principal Component Analysis (FPCA)

While smoothing and landmark registration can be considered data preprocessing, the actual statistical analysis on F_0 contours is carried out by FPCA. The input to FPCA consists of a set of (multidimensional) curves, which in our case are $(F_0(t), r(t))$ pairs, where $F_0(t)$ is represented in normalised time and $r(t)$ was defined in Eq. (1). The output is a compact model composed of a few curves, called Principal Components (PCs), which combined in appropriate amounts reproduce every input curve, with some approximation. Formally, every curve $(F_0(t), r(t))$ is decomposed as:

$$F_0(t) = m_{F_0}(t) + s_1 \cdot PC1_{F_0}(t) + s_2 \cdot PC2_{F_0}(t) + \cdots \quad (2a)$$

$$r(t) = m_r(t) + s_1 \cdot PC1_r(t) + s_2 \cdot PC2_r(t) + \cdots . \quad (2b)$$

Functions $m_{F_0}(t)$ and $m_r(t)$ are the mean curves, i.e. the curves obtained by computing the mean value of all input curves at each instant in (normalised) time in the F_0 and in the r dimension, respectively. Functions $PC1_{F_0}(t)$ and $PC1_r(t)$ are the first PC curves in the F_0 and in the r dimension, respectively, and the same holds for the second and further PCs. Finally, s_1, s_2, etc. are the so-called *PC scores*, which are coefficients that differ for every input curve and that determine the amount of correction to the mean $(m_{F_0}(t), m_r(t))$ that each PC should contribute in order to approximate a given curve $(F_0(t), r(t))$. PCs are ordered by decreasing explanatory power, in terms of percentage of explained variance. In order to avoid that the variance of one of the two dimensions dominates the other, a fixed weight is usually applied to the curves; here we have omitted that for the sake of simplicity. Crucially, Eq. (2a) and (2b) share the same PC scores. This means that PCs act jointly on the mean $(m_{F_0}(t), m_r(t))$, thus capturing variations in the shape of time-normalized contours and in their segment durations jointly (Gubian et al., 2011). All PCs are independent from one another, i.e. the shape and duration variations described by PC1 are not correlated with those described by PC2, PC3, etc. This means that given an input curve and its s_1 score, we cannot predict its s_2 and further scores.

Figure 3 provides an illustration of the mechanism of Eq. (2). The solid curves in the left panels are the mean $m_{F_0}(t)$ (top) and $m_r(t)$ (bottom), respectively. The right panels display $PC1_{F_0}(t)$ (top) and $PC1_r(t)$ (bottom), respectively. The effect of combining mean and PC curves is illustrated by the $+/-$ curves in the left panels, where Eq. (2) limited to PC1 is applied. The standard deviation of PC1 score $\sigma_{s_1} = 0.95$ is used as value for s_1 with positive or negative sign. Note that the final rise of the mean $m_{F_0}(t)$ tends to be accentuated/dumped when $\sigma_{s_1} \cdot PC1_{F_0}(t)$ is added/subtracted. The mean $m_r(t)$ is close to the value 0.0, as it represents the average of the log relative rate curves across the data set, which tend to cancel out. Adding/subtracting $\sigma_{s_1} \cdot PC1_r(t)$ to/from $m_r(t)$ makes the first and last part of the word faster/slower. Crucially, the variations in the two dimensions co-occur, e.g. a steeper final rise in F_0 is accompanied by a shorter duration of the first and last segments.

PC scores can be used as variables in ordinary statistical analyses. Once a prediction is obtained in terms of PC scores, this has to be translated back into its meaning in terms of F_0 contous. Suppose we build a model based on s_1 and we obtain a predicted value $\hat{s_1}$ from it. By substituting $s_1 = \hat{s_1}$ in Eq. (2) we obtain two predicted curves $\hat{F_0}(t)$ and $\hat{r}(t)$. While the former is of immediate interpretation, the latter is not. If we are interested in the durations of segments delimited by the landmarks, we can compute them by inverting Eq. (1). Let d_i be the normalised duration of the i-th segment, which spans the interval $[t_{i-1}, t_i]$ on the normalised time axis t. The duration $\hat{d}_i$ in the original time corresponding to the warping induced by $\hat{r}(t)$ is given by:

$$\hat{d}_i = \int_{t_{i-1}}^{t_i} \exp(-\hat{r}(t))dt. \tag{3}$$

4 Other approaches

The procedure described above is charaterised by two main aspects. One is the use of Functional PCA as a shape-to-number converter, the other is the use of Eq. (1) as a way to represent duration/rate variations across realisations. While to my limited knowledge no other comprehensive approach has been proposed for the latter, other more sophisticated methods are available as alternative for the former. In the following, first I briefly discuss advantages and disadvantages of the FPCA approach, then I refer to some alternative methods for which software packages are freely available.

The use of FPCA has two advantages. One is modularity, because FPCA acts as a sort of plug-in that substitutes only a part of the statistical modeling workflow, namely the quantitative description of curve shapes, while the rest of the analysis can be carried out without changing one's favourite modeling paradigms, e.g. linear mixed effects (LME) models. The other advantage is that FPCA provides a description of the curves that does not take into account any prior knowledge in terms of classes or factors. This provides a separation between the description of the curves' variation space and the factors that we want to characterise in terms of curve shapes. This is valuable in exploratory studies, when we are not only interested in some pre-defined factors but also into the way the data varies because of other (unknown) factors or noise. On the other hand, FPCA exposes the analysis to at least one limitation, namely that all sources of variability are described in the same space and no prior information on the structure of this space can be used. For example, building a LME model on PC scores obtained from FPCA is a sub-optimal strategy, since FPCA is not using the information on the structure of variance that random factors account for in LME models. A recent survey by Shang (2011) tells more about FPCA.

In the literature, a number of approaches that consist of a single modeling step, as opposed to two steps, i.e. FPCA + ordinary model, are available. First, functional linear models as well as functional t-test are tools included in the R `fda` package (e.g. see Cheng and Gubian (2011); Cheng et al. (2013)). More sophisticated approaches exist that allow one to build a functional LME model. R software for mixed effects FDA is introduced by Greven et al. (2010). A more comprehensive R package that extends `fda` is illustrated by Febrero-Bande and Oviedo de la Fuente (2012). Finally, a wavelet-based approach on FDA is proposed in Morris and Carroll (2006); Rausch et al. (2012).

References

Cheng, C., Chen, J.-Y., and Gubian, M. (2013). Are mandarin sandhi tone 3 and tone 2 the same or different? the results of functional data analysis.

Cheng, C. and Gubian, M. (2011). Predicting taiwan mandarin tone shapes from their duration. In *INTERSPEECH*, pages 1073–1076.

de Boor, C. (2001). *A Practical Guide to Splines, Revised Edition.* Springer, New York.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package fda.usc. *Journal of Statistical Software*, 51(4):1–28.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic journal of statistics*, 4:1022.

Gubian, M., Boves, L., and Cangemi, F. (2011). Joint analysis of f0 and speech rate with functional data analysis. In *ICASSP*, pages 4972–4975.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199.

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

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis - 2nd Ed.* Springer.

Rausch, P., Morris, J. S., Sommer, W., and Krifka, M. (2012). When you are thrown a curve: Two R packages for swerving with the wavelet-based functional mixed model.

Shang, H. L. (2011). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, pages 1–22.

Turco, G. and Gubian, M. (2012). L1 prosodic transfer and priming effects: A quantitative study on semi-spontaneous dialogues. In *Proceedings of Speech Prosody 2012*, Shanghai, China.

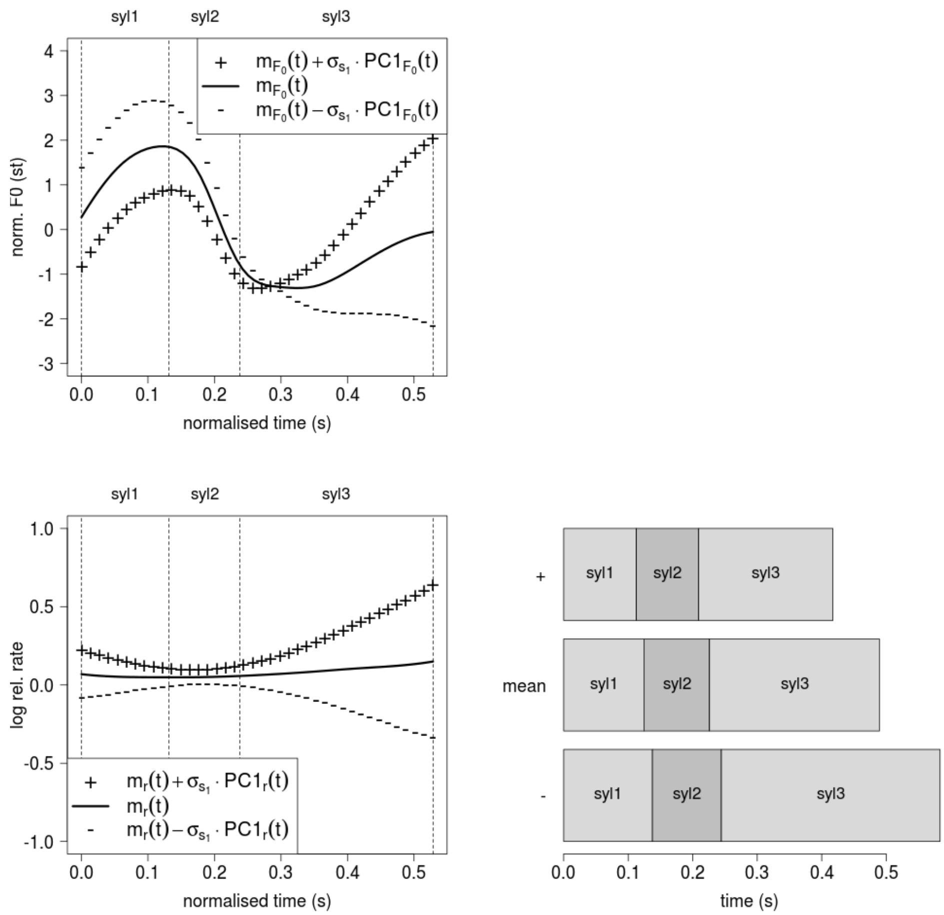

Figure 3: Illustration of Eq. (2a) (top) and (2b) (bottom), limited to PC1. The solid curves are the mean curves for the F_0 (top) and r (bottom left) dimension, respectively; $+/-$ curves correspond to Eq. (2) with $s_1 = \pm\sigma_{s_1}$. Bottom right panel: duration rappresentation obtained by applying Eq. (3) to the r curves on the left. Data from the analysis of 93 realisations of a three-syllabic word.