COGS 401 Report

Three Dimensions of Sentence Prosody-
A Functional Data Analysis Approach

Yulan Feng

July 2017

1 Introduction

Functional Data Analysis (FDA) is a relatively new technology in the field of statistical analysis [Ramsay and Silverman, 2002, Ramsay and Silverman, 2005], where samples in FDA are considered as functions instead of data points. Models based on FDA take advantage of derivative information of the functional data, and the ordering on the dimensions when by considering functional data as multivariate data [Müller et al., 2006].

Though not widely used, approaches based on FDA have recently achieved some meaningful results in linguistics studies [Parrell et al., 2013, Zellers et al., 2010]. In this project, we attempt to apply a FDA model similar to the one used in [Gubian et al., 2015] to investigate the interactions of the three dimensions of prosody sentence, and to compare the results with the ones we obtained from the previous study [Wagner and McAuliffe, 2017]. In addition, we aim to find out if these 3 features are recoverable from the F0 contours, and if there is a significant contrast.

The three dimensions include the type of speech act of an utterance (declarative v.s. interrogative intonation), the location of semantic focus (prosodic stress), and the syntactic constituent structure (phrasing).

We collected our data by having participants read a set of sentences in 16 different conditions without knowing the intention of the experiment. Take sentence (1) for example,

(1) I thought they said Marion or Marvin, and Sarah arrived. But in fact they said that **Marion** or **Marvin**, and **Nolan** arrived.

the three names in bold are our target nouns and their recorded F0 contours are preprocessed as in [Wagner and McAuliffe, 2017]. Regarding the intonation dimension, participant were asked to read the sentence in a declarative and a interrogative manner respectively; regarding the structure dimension, participants were asked to read the three names with left phrasing ([Marion or Marvin] and Nolan) and right phrasing (Marion or [Marvin and Nolan]) respectively; regarding the focus dimension, the four conditions include foci on the three names respectively in addition to a wide focus on the entire coordinate structure.

2 Experiment

In our experiment, there are mainly 3 stages: processing, smoothing, and functional data analysis. The first two stages basically serve as preparing the dataset needed for the final analysis. To be consistent with the previous experiment [Wagner and McAuliffe, 2017], we run our experiments based on 3 datasets, the original data, the residualized data, and the subset of data that only study the second word.

2.1 Processing

In addition to extracting features from the data [Wagner and McAuliffe, 2017] that will be used in the FDA, we also generate some exploratory plots in order to better understand the characteristics of the dataset. The features we extract include duration, time points for the 3 stressed vowels, speaker, relative F0 corresponding to time list(element norm time) In this project, each sample is a discourse that contains various numbers of time points with corresponding relative F0, and we have 351 discourses in total. We clean up the raw data by deleting rows with null $F0_relative$ cells, and filtering out discourses that lack enough stressed vowels.We also recalculate element normalized time to avoid duplicates so that the functions are injective.

We randomly plot 20% of raw data. As can be seen in figure 1, the raw sample contours are very messy and hard to find patterns.

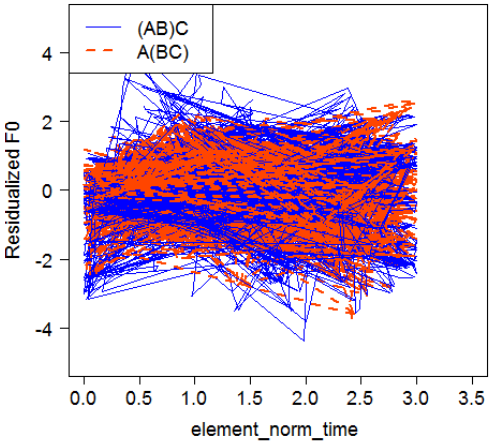
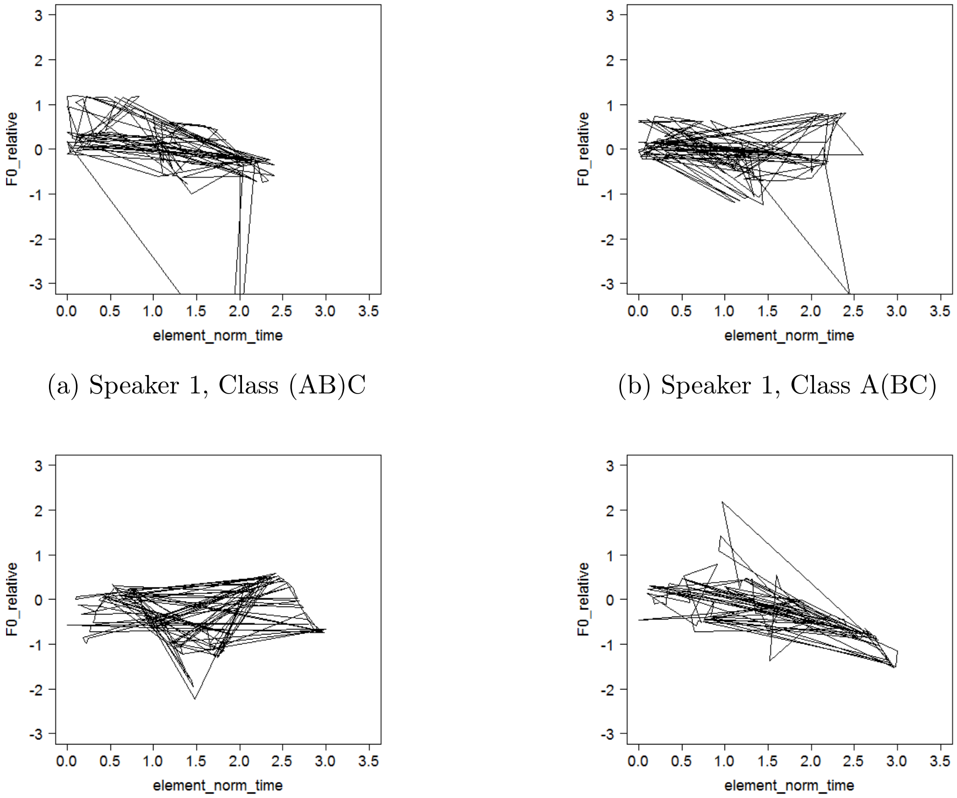


Figure 1: 20% of raw data

As we have 34 different speakers, in order to estimate the inter-speaker variability, we randomly select two speakers, and plot their relative F0 against element norm time under the same conditions except structure.(Declarative intonation, and First focus).



(a) Speaker 1, Class (AB)C

(b) Speaker 1, Class A(BC)

(c) Speaker 2, Class (AB)C

(d) Speaker 2, Class A(BC)

Figure 2: 2 speakers with 2 classes

The plots in figure 2 may be considered as *prima facie* evidence that inter-speaker variability in f0 contours is larger than class variability in this case.

2.2 Smoothing

Our raw data were recorded at discrete times. In this stage, we need to transform our sampled contours into smooth continuous functions of time in order to allow evaluation of record at any time point, evaluate rates of change, and to reduce noise. Take a simplified example where the $i-th$ sample contour is considered as

$$y_i = x(t_i) + \epsilon i,$$

where t_i is the time point, and ϵi is the error term. With smoothing, we represent $x(t)$ as

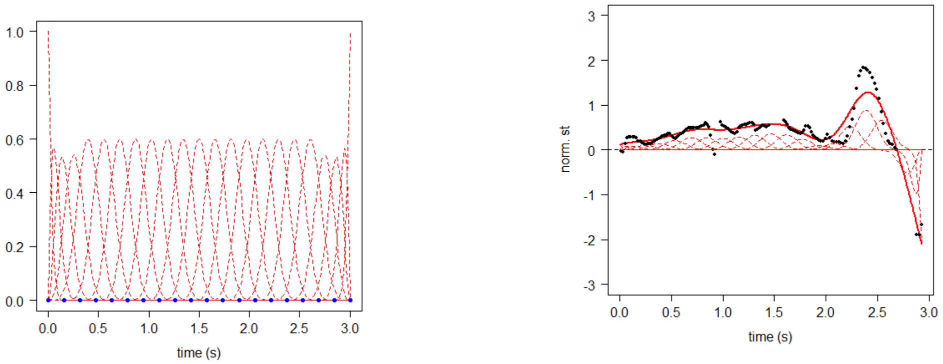
$$x(t) = \sum_{j=1}^{K} \phi(t)c,$$

where $\phi(t)$ is the basis system for x, K is the number of basis functions we choose. The coefficient c is the smoothing penalty defined as

$$c = argmin \sum_{i=1}^{n} (y_i - x(t_i))^2 + \lambda \int [Lx(t)]^2 dt,$$

where $Lx(t)$ measures "roughness" of x and λ is a regularization parameter that determines the importance of smoothness relative to fitting error. It trades-off fit to the y_i and roughness.

In our case, as F0 contours have a wide range of shapes, we adopt B-splines [De Boor, 2001] as our basis system. A B-spline is a sequence of polynomial segments joined end-to-end that, multiplied by appropriate weights and summed together, approximate a sampled data contour. This system is defined by the order m of the polynomial, and the location of the knots. We choose m=4, and set k knots equally spaced on the time axis. The value of k, together with the smoothing parameter λ, are subject to tuning, as large k with small λ will cause overfitting, and small k with large λ leads to underfitting. We use generalized cross-validation (GCV) [Ramsay and Silverman, 2005] to find a pair of (k, λ) that minimizes GCV error. The B-splines we choose to use, and an example of how sample contours are smoothed with $k = 20, \lambda = 10^{-4}$ are shown in figure 3



(a) B-spline curves

(b) A sample contour smoothed by B-splines with k=20 and $\lambda = 10^{-4}$

Figure 3: Example of smoothing

In figure 4, we plot the same subset of data as in figure 1. It is easy to observe that the data now look cleaner, and sample contours are also easier to track

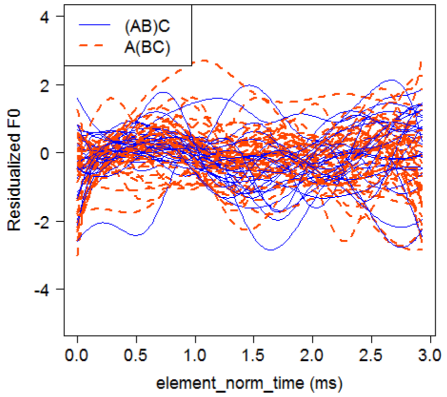
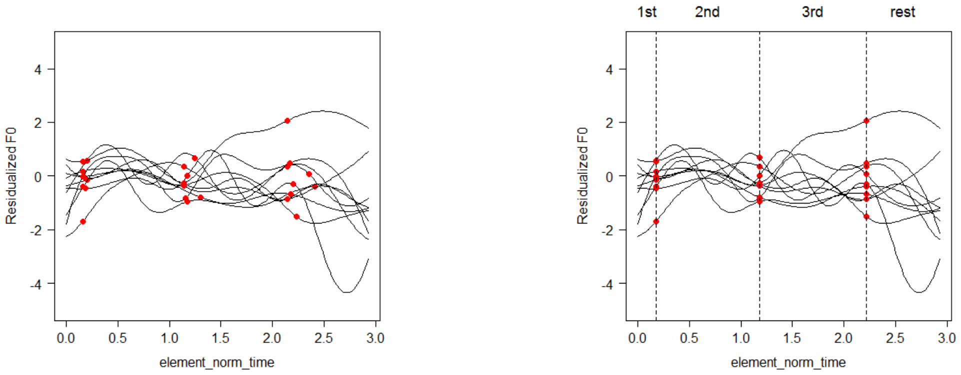


Figure 4: 20% of smoothed data

Though we are already using a normalized time system, where the raw time points were already adjusted relative to 1 within the word and in order (the 4th word of interest has all of its time points between 4 and 5, the 5th word of interest has time points between 5 and 6), as the duration of each sample varies, we still need to time-align points on different contours that correspond to the same event (stressed vowels in our case). This process is called landmark registration, where the landmarks are our stressed vowels.

To better illustrate this process, we randomly select 5% of smoothed data, and plot them both before (5a) and after (5b) landmark registration. The shapes of the smoothed curves do not change significantly, but the red points representing the stressed vowels are better aligned.



(a) 5% of data **before** landmark registration

(b) 5% of data **after** landmark registration

Figure 5: Example of landmark registration

2.3 Functional Data Analysis

In order to get a low-dimensional interpretation of our data, we apply a Functional Principal Component Analysis (FPCA) approch to our processed data. It is based on a common assumption related to the Karhunen–Loève decomposition [Fukunaga and Koontz, 1970], such that $x(t)$, after subtracting the mean function, can be represented by only the first few (usually 2 or 3) eigenfunctions:

$$X(t) - \mu(t) \approx \sum_{k=1}^{m} \xi_k \varphi_k(t),$$

where m is the number of functions, ξ_k represents the weight associated with the k-th eigenfunction (principal component) such that:

$$\xi_k = \int_T (X(t) - \mu(t))\varphi_k(t)dt$$

Therefore, FPCA provides a model of the set of input contours in terms of combinations of a small number of curves, namely the mean curve ($\mu(t)$), and the principal component (PC1,PC2…) curves, plus weights (s1,s2..) for the PC curves. $\mu(t)$ is obtained by computing the mean of all smoothed and landmark-registered f0 contours at each instant in time. The PC curves are numbered from 1 onwards and are computed by the FPCA algorithm based on the same principles as ordinary PCA [Baayen, 2008, Jackson, 2003]. The rank of the PCs reflects the decreasing percentage of variance in the input data that the PCs explain. Figure 6 shows how the fist 2 principle curves look like in the experiment with the original data. The PC curves are actually the same for all three dimensions with the original data, as data in FDA are supposed to be unlabeled.

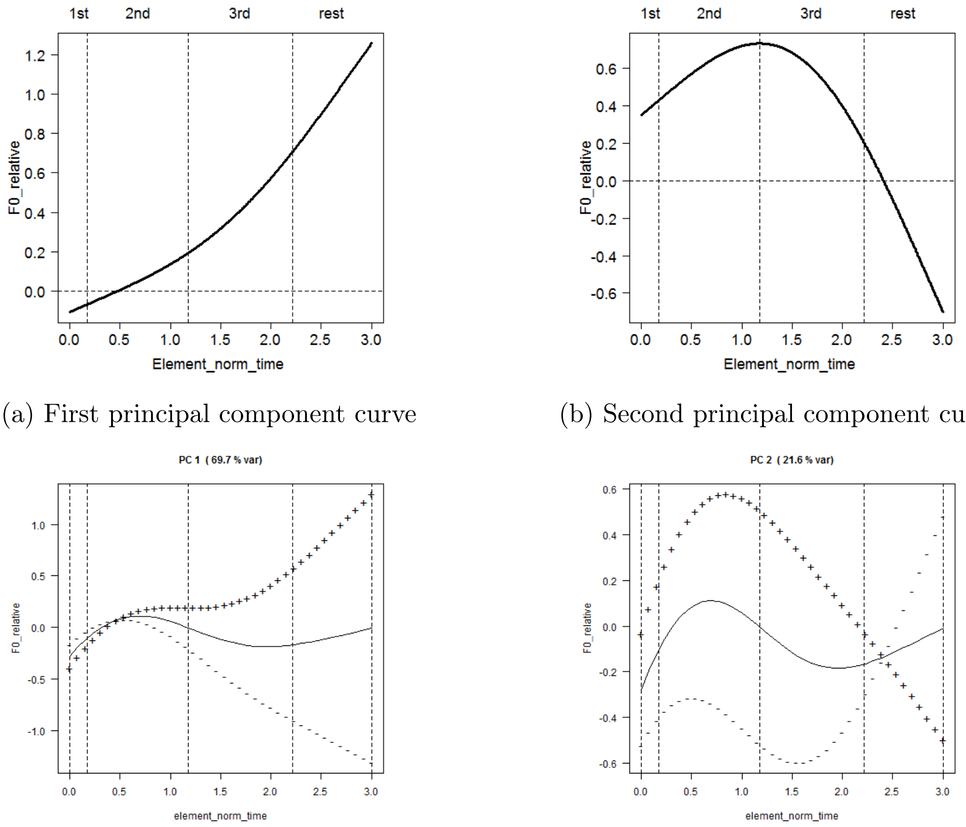
3 Results

We use 2 types of plots to estimate the contrast of different classes within each dimension: a plot of mean PC curves for each class, and plot reflecting the correlation between class and PC scores.

Figures 7 and 8 display the plots for each dimension based on the original data where all three words are included.

We also run the FPCA model on the residualized data and the subset data of the second word. The residualized data is obtained by taking the residuals from the fitted linear mixed-effects regression models, where response is the relative F0, and predictors consists of random effect for participants, and fixed effects including time, the 3 dimensions excluding the target one, and interactions of these fixed effects.

Due to limit of space, we only show the scatter plots of the structure dimension based on these 3 datasets in figure 9.



(a) First principal component curve

(b) Second principal component curve

(c) Effect of adding and subtracting the first principal component curve

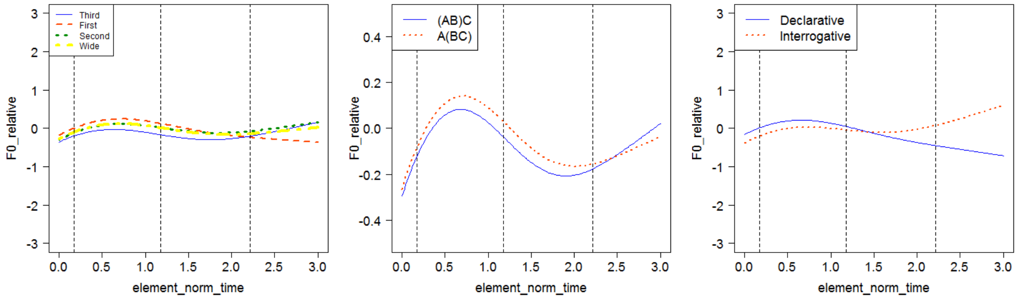
(d) Effect of adding and subtracting the second principal component curve

Figure 6: Example of principal components (PCs)

predictors	F-statistic	p-value	R^2
Intonation	64.31	<.001	.155
Focus	.507	.477	.001
Structure	.375	>.5	.001
Speaker (Intonation)	.064	>.5	.0001

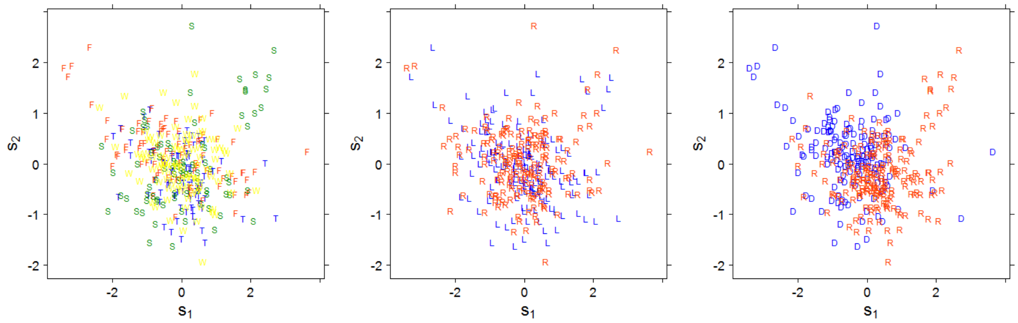
Table 1: F-test results of the 3 dimensions based on the residualized data. Speaker is tested with the residualized intonation dataset.

In addition to plotting the PC curves and scores, we also run statistical analysis using F-test to find out whether the class contrast within each dimension is statistically significant. Table 1 shows part of the results based on the residualized data. In addition, we also apply F-test to the speaker's class in order to check if there is any speaker-related effects in our



(a) Focus (b) Structure (c) Intonation

Figure 7: Class-specific mean curves



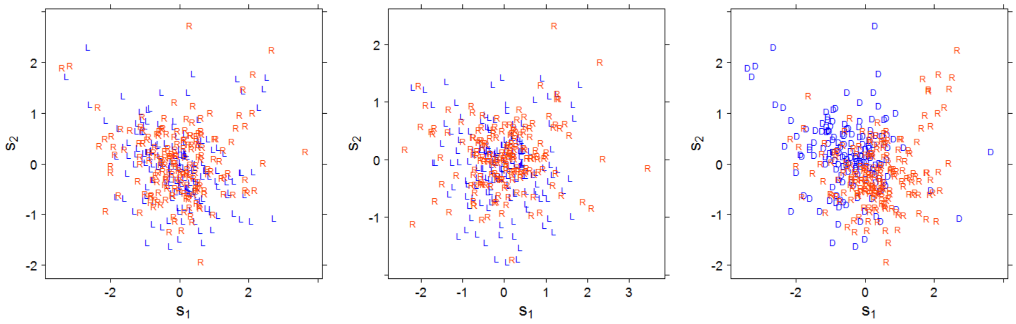
(a) Focus (b) Structure (c) Intonation

Figure 8: Scatter plots for the first 2 PC scores. In the focus plot, F,S,T,W represent "first","second","third", and "wide" foci respectively; in the structure plot, L and R represent "(AB)C" and "A(BC)" respectively; in the intonation plot, D and R represent "Declarative" and "Interrogative" respectively.

experiments.

4 Discussion

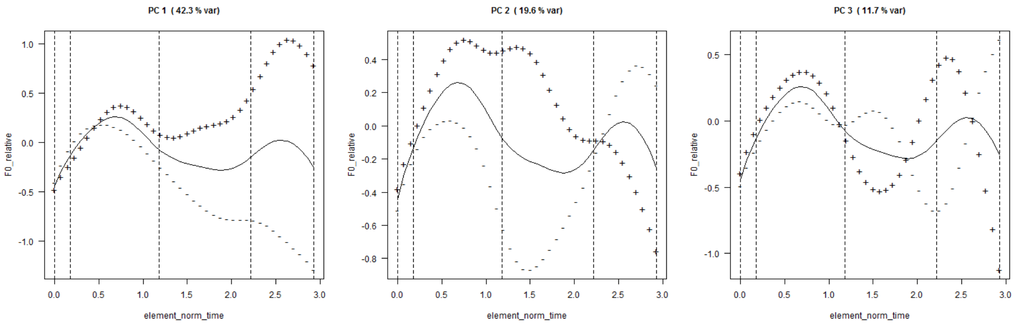
Figures 7 and 8 both suggest that only within the intonation dimension, the class contrast is obvious. For the other 2 dimensions, focus and structure, the mean curves for each class are very close, and the correlations between class and PC scores are also very weak. This is consistent with our results of F-tests. On the other hand, though our plots in figure 2 show that the speaker variability is large, the F-statistic and p-value as in 1 suggest no evidence that the speaker-related effect is significant to class contrast. Surprisingly, in contrast to our previous study [Wagner and McAuliffe, 2017], residualizing out the effects



(a) Original data (b) Residualized data (c) Subset data of word 2

Figure 9: Scatter plots of the structure dimension based on all 3 datasets

of all other dimensions does not affect the interactions among the 3 dimensions. In fact, take residualized intonation and structure data for example (figure 10 and figure 11), the shapes of and portions of variance explained by the first two PC curves almost coincide with each other. The only significant difference is the third PC, which explains a relatively small, though not negligible, portion of variance.

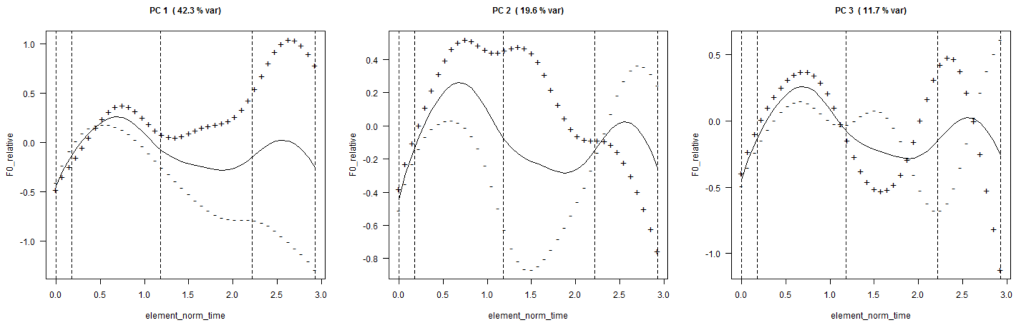


(a) PC1 (b) PC2 (c) PC3

Figure 10: Effects of adding (+) and subtracting (-) PC from mean curve on residualized intonation data

These results suggest that structure and intonation do affect each other.

In future work, we will need to conduct more empirical analyses to study why the two approaches lead to different conclusions based on the same data. In addition, we need to reconsider if models based on functional data analysis are indeed suitable for our project.



(a) PC1

(b) PC2

(c) PC3

Figure 11: Effects of adding (+) and subtracting (-) PC from mean curve on residualized structure data

References

[Baayen, 2008] Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

[De Boor, 2001] De Boor, C. (2001). *A practical guide to splines; rev. ed.* Applied mathematical sciences. Springer, Berlin.

[Fukunaga and Koontz, 1970] Fukunaga, K. and Koontz, W. L. (1970). Application of the karhunen-loeve expansion to feature selection and ordering. *IEEE Transactions on computers*, 100(4):311–318.

[Gubian et al., 2015] Gubian, M., Torreira, F., and Boves, L. (2015). Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49:16 – 40.

[Jackson, 2003] Jackson, J. E. (2003). *A Users Guide to Principal Components.* John Wiley & Sons.

[Müller et al., 2006] Müller, H.-G., Stadtmüller, U., and Yao, F. (2006). Functional variance processes. *Journal of the American Statistical Association*, 101(475):1007–1018.

[Parrell et al., 2013] Parrell, B., Lee, S., and Byrd, D. (2013). Evaluation of prosodic juncture strength using functional data analysis. *J. Phonetics*, 41(6):442–452.

[Ramsay and Silverman, 2002] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis.* Springer.

[Ramsay and Silverman, 2005] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer.

[Wagner and McAuliffe, 2017] Wagner, M. and McAuliffe, M. (2017). Three dimensions of sentence prosody and their (non-)interactions. In *Phonetics and Phonology in Europe*, Universität Köln.

[Zellers et al., 2010] Zellers, M., Gubian, M., and Post, B. (2010). Redescribing intonational categories with functional data analysis. In Kobayashi, T., Hirose, K., and Nakamura, S., editors, *INTERSPEECH*, pages 1141–1144. ISCA.