Journal of Phonetics 49 (2015) 16–40

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Technical Article

Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts

Michele Gubian [a,*], Francisco Torreira [a,b], Lou Boves [a]

[a] Centre for Language and Speech Technology, Radboud University Nijmegen, Erasmusplein 1, 6525 HT, Nijmegen, The Netherlands
[b] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

A R T I C L E I N F O

Article history:
Received 30 December 2013
Received in revised form
27 September 2014
Accepted 2 October 2014

Keywords:
Functional Data Analysis
Cue Trading
Dynamic trajectories
Diphthong and hiatus
European Spanish

A B S T R A C T

The study of phonetic contrasts and related phenomena, e.g. inter- and intra-speaker variability, often requires to analyse data in the form of measured time series, like f_0 contours and formant trajectories. As a consequence, the investigator has to find suitable ways to reduce the raw and abundant numerical information contained in a bundle of time series into a small but sufficient set of numerical descriptors of their shape. This approach requires one to decide in advance which dynamic traits to include in the analysis and which not. For example, a rising pitch gesture may be represented by its duration and slope, hence reducing it to a straight segment, or by a richer coding specifying also whether (and how much) the rising contour is concave or convex, the latter being irrelevant in some context but crucial in others. Decisions become even more complex when a phenomenon is described by a multidimensional time series, e.g. by the first two formants.

In this paper we introduce a methodology based on Functional Data Analysis (FDA) that allows the investigator to delegate most of the decisions involved in the quantitative description of multidimensional time series to the data themselves. FDA produces a data-driven parametrisation of the main shape traits present in the data that is visually interpretable, in the same way as slopes or peak heights are. These output parameters are numbers that are amenable to ordinary statistical analysis, e.g. linear (mixed effects) models. FDA is also able to capture correlations among different dimensions of a time series, e.g. between formants F_1 and F_2. We present FDA by means of an extended case study on diphthong – hiatus distinction in Spanish, a contrast that involves duration, formant trajectories and pitch contours.

1. Introduction

As early as the 1970s it has been shown that the same phonetic phenomenon can be related to several different acoustic features (Repp, 1981; Slis & Cohen, 1969). Also, the information carried by some features, such as pitch or formants, is encoded in dynamical changes over time, rather than as fixed-length sequences of scalars that reflect the feature value at crucial points, such as the beginning and the end of a segment. Still, for perfectly understandable and legitimate reasons, most phonetic research has been based on the analysis of individual features represented with a small number of scalar values (e.g. formant values in the centre of vowels, minimum or maximum f_0 values in Hz or semitones, alignment of f_0 minima or maxima relative to the beginning of a segment). Probably, phoneticians prefer this approach because of technical and methodological constraints. For instance, conventional statistical methods require that all observations are expressed as a fixed-length sequence of numbers, and only one dependent variable can be investigated at a time. Such an approach requires phoneticians to decide in advance the points in time where a feature value is obtained, and therefore entails the risk of ignoring potentially relevant detail. For example, a rising pitch movement may be represented by its duration and slope, which effectively reduces the trajectory to a straight line, or it may be represented by a

* Corresponding author. Present address: School of Experimental Psychology, University of Bristol, office 3D3, The Priory Road Complex, Priory Road, Clifton BS8 1TU, UK. Tel.: +44 117 3317893.
E-mail addresses: m.gubian@let.ru.nl, mm14722@bristol.ac.uk (M. Gubian), francisco.torreira@mpi.nl (F. Torreira), l.boves@let.ru.nl (L. Boves).
URL: http://lands.let.ru.nl/FDA/ (M. Gubian).

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.10.001

more complex coding that captures the concavity or convexity of the trajectory. Such details in the f_0 shape may be irrelevant in some context, but crucial in others (Dombrowski & Niebuhr, 2010). An alternative method to code pitch, formant or intensity contours consists in using the coefficients from a polynomial fit (Andruski & Costello, 2004; Grabe, Kochanski, & Coleman, 2007). This may be a good solution for short curves with a single extremum, but it leaves us with the burden of interpreting the coefficients of the polynomial when fitting more complex trajectories.

In this paper, we introduce *Functional Data Analysis* (FDA; Ramsay & Silverman, 2005) as a new method for analysing phonetic phenomena involving dynamic changes across multiple acoustic parameters. In this sense, this paper extends previous presentations of FDA as a tool for phonetic analysis (Cheng, Xu, & Gubian, 2010; Gubian, Boves, & Cangemi, 2011; Zellers, Gubian, & Post, 2010). We show that FDA allows us to apply familiar statistical methods (e.g. linear regression, principal component analysis) to dynamic features by representing them as continuous functions, and that this can be done for multiple features in a single joint analysis. To illustrate this, we revisit the diphthong–hiatus distinction in Spanish, a contrast in which several phonetic features have been implicated (duration, formant trajectories, f_0 alignment, see Aguilar, 1999; Hualde & Prieto, 2002; Torreira, 2007). In particular, we show how FDA can be used to reveal the trading relations between these features, and, more specifically, how different speakers trade the features in different ways.

The remainder of the paper is structured as follows. In Section 2, we describe the diphthong–hiatus distinction in Spanish, and the data that will be used in the following sections to illustrate the powers of FDA. Section 3 contains a practical introduction to using FDA for phonetic research. In Section 4 we first investigate the role of duration, formant trajectories and f_0 alignment in the diphthong–hiatus contrast by treating each feature separately. After this, we present a joint analysis of the three features, and show that features that seem to play a powerful role when observed in isolation may not be as important when other features are taken into account. Section 5 contains a discussion of the advantages and limitations of FDA in relation to traditional analysis, with particular attention on the role of prior knowledge, and a sketch of the opportunities offered by FDA to the study of different kinds of phonetic phenomena. Finally, Section 6 concludes the paper. In conjunction with the Appendices, the Supplementary materials[1,2] and the web site on FDA maintained by the first author,[3] this paper should enable researchers to apply FDA to their own data.

2. Case study

In Spanish, vowel sequences of rising sonority (e.g. /ie/, /ia/, /ua/) are said to be syllabified in a generally predictable manner, depending on the location of lexical stress. Hiatuses (i.e. /Ci.a/) occur when lexical stress is on the initial high vowel; otherwise, the vowel sequence is realised as a diphthong /Cja/. Despite the fair degree of generality achieved by this rule, some lexical exceptions have been noted that make Spanish a language with a phonological contrast between hiatuses and diphthongs (e.g. *diente* [’djen.te] ‘tooth’ vs. *cliente* [cli.’ente] ’client’; *italiano* [i.ta.lja.no] ‘Italian’ vs. *liana* [li.’a.na] ‘liana’).

Although the idea of a phonological contrast between diphthongs and hiatuses in Spanish (*D*/*H* contrast from now on) is not controversial, its distribution in the lexicon, and the consistency with which it is realised phonetically, appear to vary across dialects and speakers. Hualde (2005) mentions that hiatuses are much more common in Castilian Spanish than in Latin American dialects, and also points to the existence of idiolectal variation within dialects. Several phonetic studies have investigated the acoustic basis of this contrast in several varieties of Spanish. In Aguilar (1999), rising diphthongs and hiatuses were extracted from a Barcelona Spanish corpus of map-task conversations and read sentences. The durations of the vowel sequences were measured, and formants were modeled with second-order polynomials capturing the slope and curvature of their trajectories. It was found that hiatuses have a longer average duration and a greater degree of curvature in the F_2 trajectory than diphthongs, both in conversational and in read speech. Hualde and Prieto (2002) investigated the *D*/*H* contrast by asking Madrid Spanish speakers to syllabify and read series of words containing the vowel sequence *ia* and by measuring the duration of the produced vowel sequences. They found that, as reported in Aguilar (1999), speakers produced longer vowel sequences in cases categorised as hiatuses than in those categorised as diphthongs. However, they also found that the duration distributions of the diphthong and hiatus groups overlapped considerably in the case of some speakers. More recently, Torreira (2007) analyzed the alignment of rising pitch accents in Spanish segmental sequences involving a similar gestural content but differing in syllabic structure, including the diphthongs and hiatuses investigated in the present study. In agreement with the syllabification patterns proposed by Hualde and Prieto (2002), it was found that rising pitch accents in /ia/ hiatuses were aligned with the second vowel of the sequence, while in diphthongs, the start of rising accents was aligned earlier, presumably at the onset of the syllable containing the diphthong. Since the onset of rising pitch accents in Spanish have been reported to be aligned with the beginning of lexically stressed syllables (Prieto, van Santen, & Hirschberg, 1995; Prieto & Torreira, 2007), this study concluded that the differences in f_0 alignment must be due to differences in syllabification of /ia/ vowel sequences (i.e. /Cja/ vs. /Ci.a/). Although intonational features are suprasegmental by definition, it is possible that their alignment with segmental features can be used as cues to the identity of the latter. For this reason, we will also consider f_0 alignment as a potential cue of the *D*/*H* contrast, along more straightforward vocalic features such as duration and formant trajectories.

[1] Code, data and plots not included in the text are available for download from this repository: https://github.com/uasolo/FDA-DH/. Direct link to zip bundle: https://github.com/uasolo/FDA-DH/archive/master.zip

[2] Direct link to Additional Material pdf: https://github.com/uasolo/FDA-DH/raw/master/paper/FDA-DH-Additional_Material.pdf

[3] http://lands.let.ru.nl/FDA