Supplementary material for the paper
The statistical analysis of acoustic phonetic data: exploring differences between spoken Romance languages.

Davide Pigoli
Department of Mathematics, King's College London
Pantelis Z. Hadjipantelis
Department of Statistics, UC Davis
John S. Coleman
Phonetics Laboratory, University of Oxford
John A. D. Aston†
Statistical Laboratory, DPMMS, University of Cambridge

Robustness of the analysis to preprocessing choices

The analysis of acoustic phonetic data needs a few preprocessing step to make cross-words and cross-language comparisons feasible, as detailed in the paper. However, this may raise the question of how much the results are robust with respect to specific pre-processing choices. In particular, the smoothing step and the time registration step can be performed with different techniques and parameter choices, each with advantages and disadvantages. Here we replicate the analysis of the paper changing the methods used in these preprocessing steps. This highlights also the flexibility of the proposed approach, where practitioners can choose the most appropriate (or personally favourite) method at each stage of the procedure.

For what concerns the smoothing step, in place of the automated robust algorithm for two-dimensional gridded data described in Garcia (2010), we rely here on thin-plate regression splines implemented in the R package `mgcv` (Wood, 2003, 2006).

Then, we standardize the time scale so that each signal goes from 0 to 1 and we select the time registration by applying the Fisher-Rao metric minimization procedure (Srivastava et al., 2011; Wu and Srivastava, 2014) to the set of curves $S_{ik}^{L}(500\text{Hz}, t)$ for each language L and digit i. Figure 1 shows the smoothed and aligned log-spectrogram for a French speaker uttering the word "un" (one) obtained with the above procedure, compared with what would have been obtained through the preprocessing described in the paper. As it can be seen, the former is less smooth but of course this can be changed by decreasing the number of thin-plate basis functions used in the smoothing.

Tables 1, 2 and 3 report the p-values of the permutation tests described in the paper for the equality of the mean, the frequency covariance functions and the time covariance

†Address for correspondence: Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, email: `j.aston@statslab.cam.ac.uk`

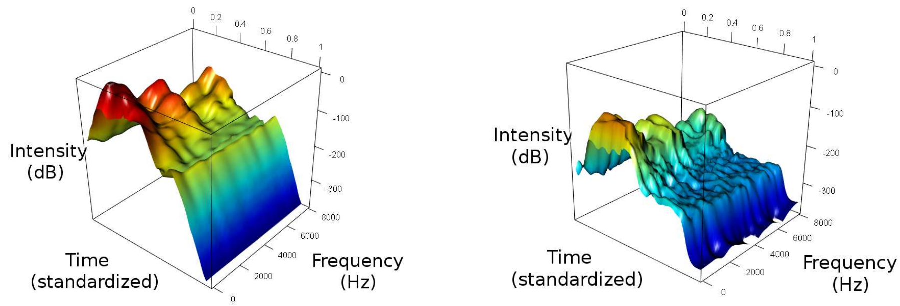

Figure 1. Left: Smoothed and aligned log-spectrogram for a French speaker pronouncing the word "un" (one), using thin-plate spline smoothing and Fisher-Rao metric minimization alignment. Right: Smoothed and aligned log-spectrogram of the same sound obtained with the preprocessing method used in the paper.

Table 1. P-values of the permutation tests for H0: $\mu_1^L = \mu_2^L = \cdots = \mu_{10}^L$ *vs* H1: at least one is different, where μ_i^L is the mean log-spectrogram for the language L and word i, for the five Romance languages.

Language	French	Italian	Portuguese	American Spanish	Castilian Spanish
$p - value$	<0.001	0.014	0.962	<0.001	0.002

functions, respectively, between the different words of each language. The results appear again in support of the hypothesis that the means are different across the words while there is no evidence of differences between the covariances. The only relevant difference appears in the test for the mean log-spectrograms across words in Castilian Spanish, where no significant difference were found with the original preprocessing while with the new preprocessing we get a p-value of 0.002. This may be be due to the fact that the Fisher-Rao metric registration is able to further reduce the variability in the sample.

Figure 2 shows colormaps of the projection of the log-spectrogram from French language to Portuguese language for digit 1 and the closest Portuguese log-spectrograms. The audio file `F2P_proj_digit5_SuppMat.wav` in the supplementary material contains the reconstructed sound for the path $S_{5,1}^{Fr\to P}(x)$, $x = (0, 1)$, connecting the spoken word *cinq* ("five") uttered by a French speaker with its projection into the Portuguese language.

Table 2. P-values of the permutation tests for H0: $C_{\omega,1}^L = C_{\omega,2}^L = \cdots = C_{\omega,10}^L$ *vs* H1: at least one is different, where $C_{\omega,i}^L$ is the marginal frequency covariance operator for the language L and word i, for the five Romance languages. The Procrustes distance is used for the test statistic.

Language	French	Italian	Portuguese	American Spanish	Castilian Spanish
$p - value$	0.38	1	0.95	0.83	1

Table 3. P-values of the permutation tests for H0: $C^L_{t,1} = C^L_{t,2} = \cdots = C^L_{t,10}$ *vs* H1: at least one is different, where $C^L_{t,i}$ is the marginal time covariance operator for the language L and word i, for the five Romance languages. The Procrustes distance is used for the test statistic.

Language	French	Italian	Portuguese	American Spanish	Castilian Spanish
$p - value$	0.91	1	0.74	0.71	0.20

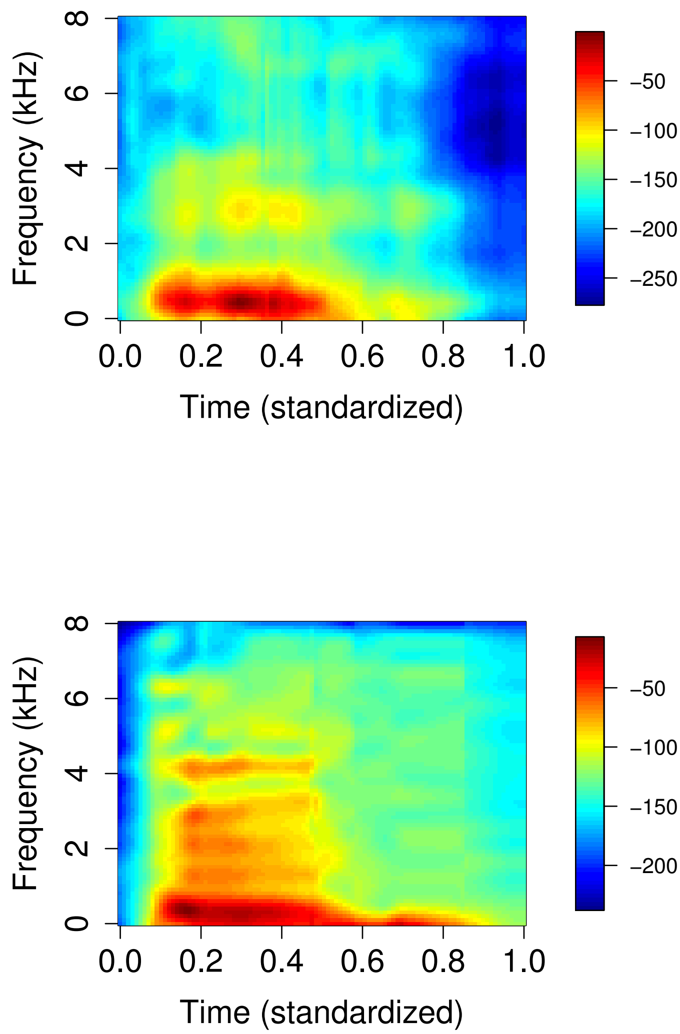

Figure 2. Colormaps of the projection of the log-spectrogram of a French speaker pronouncing the word “un” into Portuguese (top) and closest log-spectrogram of Portuguese speaker pronouncing the word “um” (bottom).

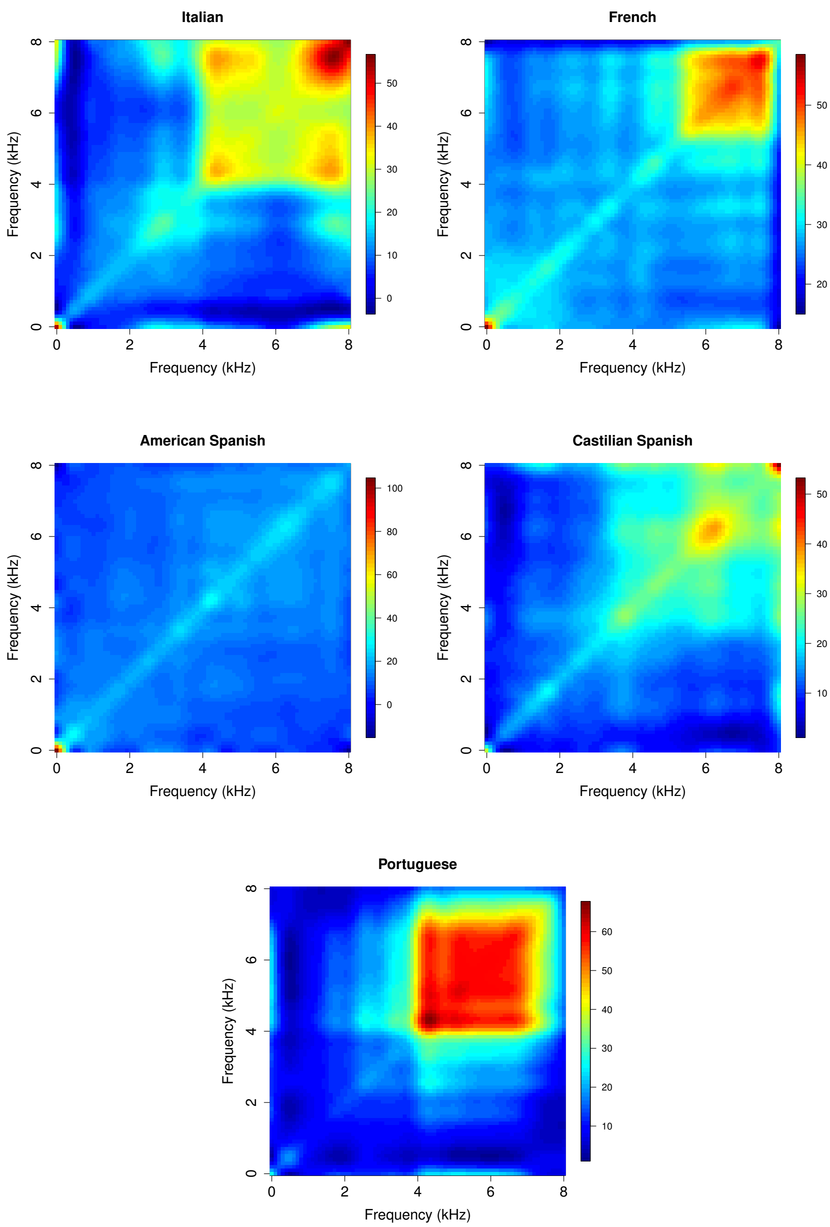

Figure 3. Colormaps of the marginal covariance function between frequencies for the five Romance languages.

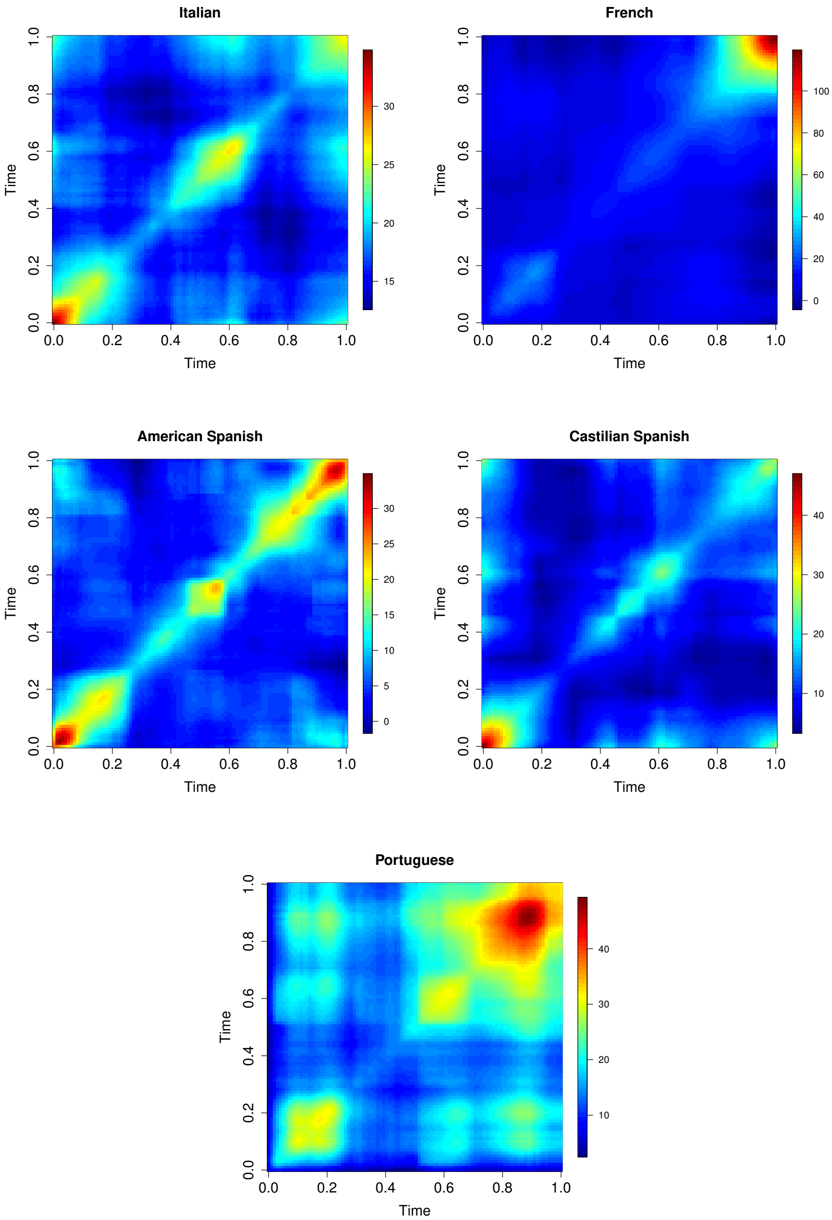

Figure 4. Colormaps of the marginal covariance function between times for the five Romance languages.

References

Garcia, D. (2010) "Robust smoothing of gridded data in one and higher dimensions with missing values", *Computational Statistics & Data Analysis*, **54**, 1167–1178.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011) "Registration of functional data using the Fisher-Rao metric", Technical report, Florida State University. arXiv:1103.3817v2.

Wood, S. N. (2003) "Thin plate regression splines", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 95–114.

Wood, S. (2006) *Generalized additive models: an introduction with R*, CRC press.

Wu, W., and Srivastava, A. (2014) "Analysis of spike train data: Alignment and comparisons using the extended Fisher-Rao metric", *Electronic Journal of Statistics*, **8**, 1776–1785.