Parametric representations of formant trajectories in Spanish vocalic sequences for likelihood-ratio-based Forensic Voice Comparison

Eugenia San Segundo Fernández

Phonetics Lab, CSIC

Introduction

It seems that from dynamic spectral properties (e.g. diphthongal formant trajectories) [1-3] we can get more information about a speaker's idiosyncrasies than, for example, from central formant frequencies, which are more constrained by the linguistic system shared by speakers of the same variety.

Objectives

To which extent are formant trajectories of Spanish vocalic sequences useful forensic parameters?

- Do some parametric representations allow better performance than others?
- Does considering together both diphthong [ja] and hiatus [ia] allow better performance than considering them separately?
- Do some Spanish vocalic sequences allow better perfomance than others?

Materials & Methods

Data: Spontaneous speech recordings from 29 adult male speakers of Castilian Spanish selected from the Ahumada Corpus [4]. Two sessions per speaker.

Acoustic analysis: Vocalic sequences [we], [je], [ja] and [ia] were manually marked [5]. For each one the trajectories of F1-F3 were tracked [6].

Curve fitting: Parametric curves (cubic polynomials and DCT, discrete cosine transforms) were fitted to the three formant trajectories.

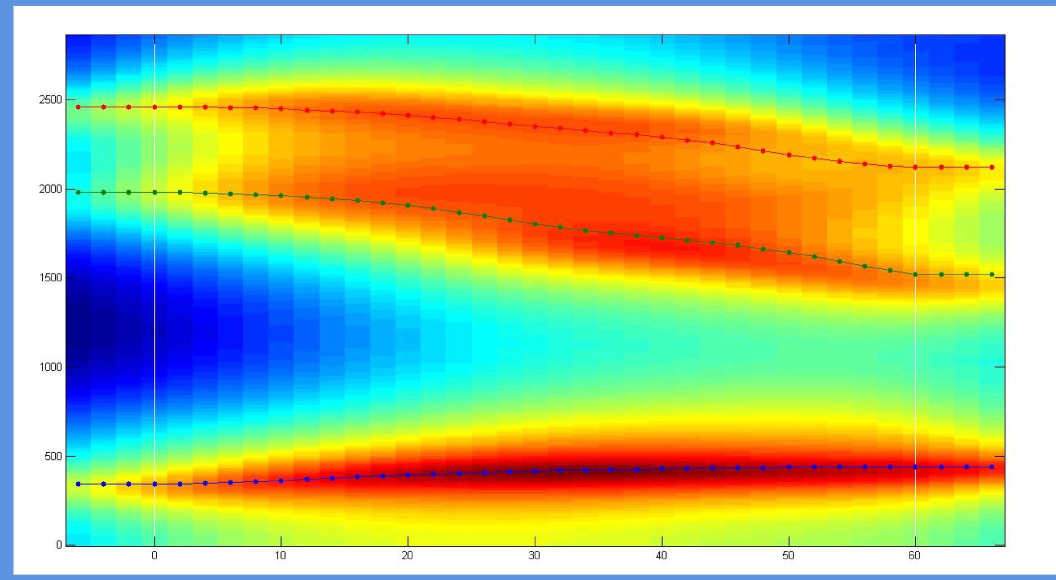

Figure 1. Formant trajectories of [je] in speaker 12.

LR calculation: Cross-validated likelihood ratios were calculated using the multivariate kernel density formula described in [7] and implemented in [8].The variables entered into the formula were the coefficient values from the parametric curve fitting.

Results

Table 1 shows the log-likelihood-ratio cost (C_{llr}) calculated for the likelihood ratios (LRs) obtained. The measure of accuracy C_{llr} [9] is a continuous function which is small for correct LRs and asymptotes towards zero as correct LRs diverge from 1, but which is large for incorrect LRs and becomes exponentially large as incorrect LRs diverge from 1. The lower the C_{llr}, the better the performance of the system.

Tippett plots in figures 2 and 3 show the results of running some of our LR systems.

	[we]			[je]			[ja]			[ia]			[ja] & [ia]		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
DCT	1.273	1.263	1.293	1.250	1.266	1.233	1.211	1.252	1.259	1.298	1.327	1.308	1.320	1.355	1.265
POLYNOMIAL	1.269	1.239	1.257	1.229	1.270	1.287	1.405	1.388	1.176	1.273	1.483	1.356	1.277	1.276	1.222

Table 1. C_{llr} results for all the formants and vocalic sequences.

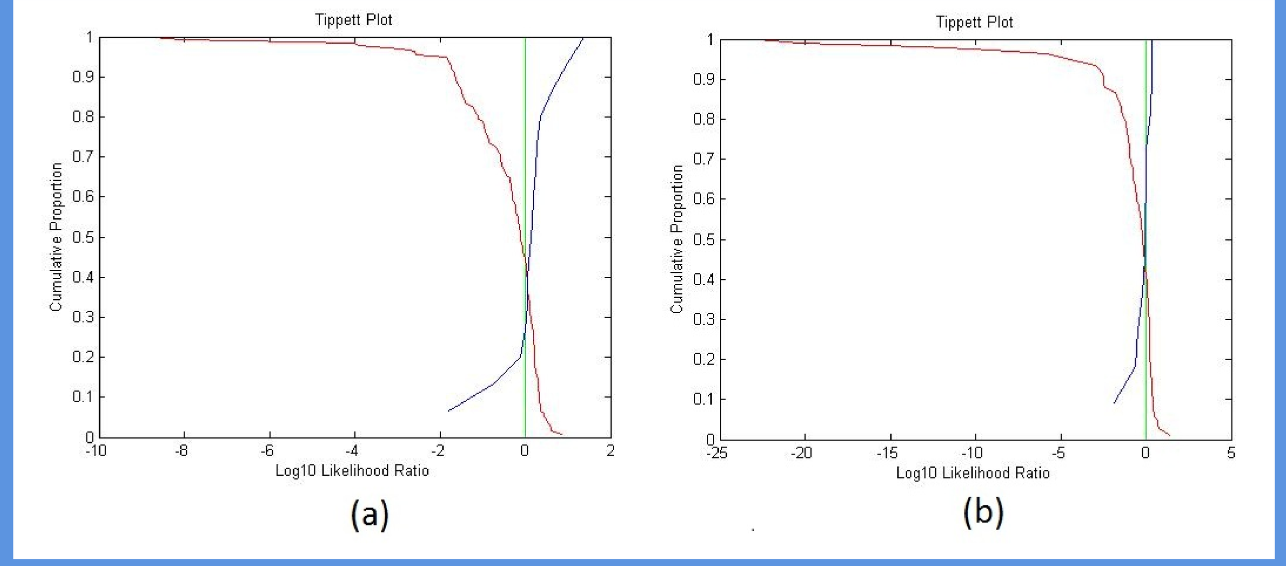

Figure 2.Tippett plots of log-likelihood ratios obtained for polynomial representations: (a) lowest C_{llr} (F3 of [ja]) and (b) largest C_{llr} (F2 of [ia]).

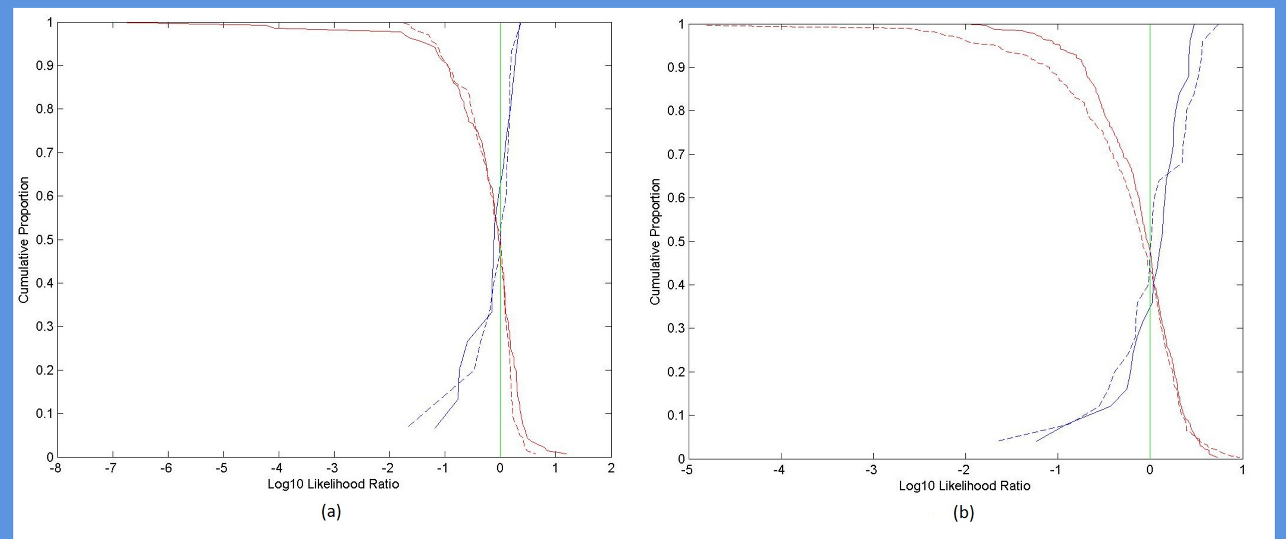

Figure. 3. Tippett plots of log likelihood ratios obtained for the formant and vocalic sequence with (a) the largest and (b) the smallest difference in C_{llr} between the DCT representation (dashed line) and cubic polynomial representation (solid line): (a) F1 of [ja] and (b) F2 of [je].

Conclusions

- There appears to be relatively little difference between using polynomial and DCT curves.
- Performance improves if we consider [ja] and [ia] separately.
- We could tell if some formants in a vocalic sequence have better performance than others but we cannot know if certain vocalic sequences perform better than others.

Further research

- Use of GMM-UBM to calculate LRs.
- Use of fusion and calibration to know if some vocalic sequences allow better performance than others.

Acknowledgements

This research was supported by a grant from the Spanish Ministry of Education. I am also indebted to Dr Geoffrey S. Morrison for providing support and guidance.

References

[1] Rose, P., Kinoshita, Y.& Alderman, T. (2006)
[2] Morrison, G.S. (2009)
[3] Enzinger, W. (2010)
[4] Ortega, J., González, J. & Marrero, V. (2000)
[5] Morrison, G.S. (2010a)
[6] Morrison, G.S. (2010b)
[7] Aitken, C.G.G. & Lucy, D. (2004)
[8] Morrison, G.S. (2007)
[9] Brümmer, N. & du Preez, J. (2006)