This paper has been accepted for publication.

Citation: Elvira-García, W.; Balocco, S., Roseano, P.; Fernández Planas, A. M. (2018). «ProDis: A dialectometric tool for acoustic prosodic data», *Speech Communication.*

DOI: https://doi.org/10.1016/j.specom.2017.12.013

Official link to publication: to be published

ProDis: A dialectometric tool for acoustic prosodic data

Wendy Elvira-García[a], Simone Balocco[b,c], Paolo Roseano[a,d], Ana Ma. Fernández-Planas[a]

[a] Laboratori de Fonètica, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain

[b] Dept. Mathematics and Informatics, University of Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain

[c] Computer Vision Center, 08193 Bellaterra, Spain

[d] Dept. of Linguistics and Modern Languages, University of South Africa, Muckleneuk campus, Theo van Wijk 9-88, 0002 Pretoria, South Africa

Abstract

This paper introduces a new tool that performs dialectometric analysis using prosodic acoustic data. This method differs from previous approaches of prosodic dialectometry in the following aspects: 1) the approach does not need a previous intonational transcription, 2) it considers all prosodic parameters (intonation, duration and intensity), 3) it offers a complete set of transparent statistical tests, 4) it is user-friendly, and fully automatic and the computation does not require a concatenation of scripts and applications. The paper also presents the results obtained using the prosody of several Romance varieties. The high correlation among the computed acoustic distances with the distances perceived by human listeners ($r= 0.8567$, $p<.0001$) shows the reliability of the approach.

Keywords

Dialectometry; prosodic data; AMPER; linguistic atlases; ProDis

1 Introduction

Dialectometry is a set of computational techniques used to classify and to investigate relationships between dialectal areas. It was first used to measure the distance among dialects of Gascon Occitan (Séguy, 1971). Data used in dialectometric analyses usually consist in phonological transcriptions of words as uttered in different survey points. They are therefore nominal data.

One of the main shortcomings due to the use of phonological segmental transcriptions is that they may not be objective. In fact, any phonological transcription implies a subjective interpretation of the acoustic data that depends on the perception and phonological knowledge of the transcriber. In order to overcome this limitation Heeringa et al. (2009) proposed to measure segmental dialectal differences directly from the numeric acoustic data of the recordings without undergoing a manual transcription stage. In their study they applied this technique to vowels using formant values and proved that dialectometric distances can be computed successfully using acoustic data. In addition, they found out that there was a high correlation between the acoustic distance and the perceived distance between dialects.

Existing dialectometric studies dealing with suprasegmental data (i.e. prosody and especially intonation) are exposed to the same risk that Heeringa et al. (2009) envisaged for segmental data: distances are usually computed using either phonological transcriptions given by the researchers, or the perceived distance expressed by the speakers of the language (Gooskens, 2002; Gooskens and Heeringa, 2006, 2004). Hence, the distances are based on subjective nominal data.

The hypothesis of this paper is that that the methodological approach successfully proposed by (Heeringa et al., 2009) for segmental data can be extended to suprasegmental data (Fernández Planas et al., 2015; Roseano et al., 2015). In other words, we argue that prosodic transcriptions are not necessary to perform dialectometry. We will show 1) that dialectal classifications can be achieved by using acoustic numeric data, 2) and that such classifications are comparable with the distances perceived by native speakers.

In order to compute and analyze the acoustic distances between the intonations of different locales a computational framework called ProDis (acronym for Prosodic Distances) was designed.

The paper is organized as follows: Section 1 states the motivation, state of the art and novelty of the project. Section 2 contains details about the method, paying special attention to materials, metric used for computing the correlation between sentences computation and creation of the similarity matrices. Section 3 summarizes the results achieved with the software applied to data from several Romance varieties and compares the results obtained by ProDis with the distances perceived by native speakers. The final section states some concluding remarks.

1.1 Motivation

Prosodic differences are a salient feature in Romance dialectal variation (Martínez Celdrán and Fernández Planas, 2003; Prieto et al., 2010). Despite being of great interest, prosody has been largely neglected in studies about dialectal variation.

Traditional dialectological descriptions relied on lexical and segmental-phonological data stored in linguistic atlases that had been published since the early 20th century. Since the last quarter of the same century, dialectometric tools have been created to analyze the lexical, morphological and segmental phonetic data contained in such atlases. Prosodic data were not taken into consideration neither in traditional dialectological studies nor in modern dialectometric approaches for two very good reasons. One side, in fact, until the last decade of the 20th century no linguistic atlas contained prosodic data. On the other side, there was no tool capable of dialectometrizing numeric acoustic prosodic data (Roseano, 2016).

This double limitation has been tackled by researchers by collecting prosodic data. In the last two decades, in fact, prosodic data has been stored in digitalized atlases that provide the necessary data for dialectometric studies, like the *Atlas Multimédia Prosodique de l'Espace Roman* (AMPER) or the *Interactive Atlas of Romance Intonation* (IARI) (Contini and Romano, 2002; Martínez Celdrán and Fernández Planas, 2003; Prieto et al., 2010). However, due to the lack of adequate tools, such data has not been fully exploited yet. In fact, the prosodic dialectometric tools existing before ProDis had limitations. Some of them, for example, were able to analyze phonological transcriptions of intonation (and, thus, were exposed to the risks pointed out by Heeringa et al. (2009)), only. Some others required to concatenate different software and scripts in complex pipelines in order to analyze prosodic

numeric data and, in addition, did not provide all usual dialectometric outputs (Fernández-Planas et al., 2015b; Moutinho and Coimbra, 2011; Roseano et al., sent). See Section 1.2 for further details on existing tools.

It is therefore clear that a suitable software is needed in order to dialectometrize the prosodic data contained in linguistic atlases.

1.2 State of the art

The creation of prosodic atlases like AMPER or IARI described in Section 1.1 provided the databases needed for dialectometric studies in the suprasegmental domain. Nevertheless, the development of tools capable of dialectometrizing such prosodic data is far less advanced. Existing attempts of dialectometric studies with prosodic data may be divided in two branches: studies with nominal data (i.e. with transcriptions of intonational contours) and studies with numeric data (i.e with values in Hertz –Hz–, semitones –st –, milliseconds –ms– and decibels –dB–). Different methods and tools have been used in each branch.

As far as the nominal branch is concerned, one has to bear in mind that since the Seventies, multiple algorithms have been created for dealing with nominal data and to calculate distances between phonological transcriptions, lexemes, and even syntax. Most of the literature regarding dialectometry since (Kessler, 1995) uses Levenshtein distances or variations of the Levenshtein algorithm. These algorithms were used to compute distances also in the prosodic domain (Gooskens, 2002; Gooskens and Heeringa, 2006). Such algorithms, in fact, have been applied to nominal prosodic data by means of different computational tools, like Gabmap (Nerbonne et al., 2011), Diatech (Aurrekoetxea et al., 2013) and Visual DialectoMetry (Haimerl, 2006), which were not specifically designed for prosodic dialectometry, and therefore often required extra work to make the analysis possible. For example (Prieto and Cabré, 2013) used Gabmap, which required a kind of transliteration of the transcriptions because it did not accept the symbols commonly used in prosodic transcriptions. On the other hand, (Fernández-Planas et al., 2015a) used a pipeline of different software consisting in an ad-hoc transcriber (called AMPER-Eti, see (Roseano and Fernández Planas, 2013), Praat scripts, Excel and SPSS. On the whole, thus, one can say that there is, to date, no single specific software capable of carrying out an easy and straightforward analysis of nominal prosodic data. In addition, another limitation of such studies is that the contours were classified using pre-determined phonological transcription. For this reason the results were subject to the user subjectivity, so discrimination between dialects is more accurate when phonetic data is exploited than using transcriptions (Heeringa et al., 2009).

As far as the numeric branch is concerned, the situation is similar, insofar as there is still no single specific software capable of carrying out an easy and straightforward analysis of numeric prosodic data. The lack of such specific software has been a frequent concern for the AMPER project (Contini and Romano, 2002) and there have been a few attempts to create an effective, simple and reliable tool for numeric dialectometric analysis. The first attempt in this direction was the program called *Stat-distances* (Moutinho and Coimbra, 2011), a discontinued software that has received criticism for not carrying out dialectometry transparently (for example, the program did not allow to see the distance matrices nor to check the parameters that it used for creating the dendrograms, see (Roseano, 2016). Another attempt in the same direction was *Calcu-Dista*, a conjunction of Praat scripts, Excel and SPSS (Fernández-Planas et al., 2015a; Roseano et al., sent). Furthermore, while this investigation is being conducted, another dialectometry method is being developed to analyze Portuguese and Galician data (Calvo and Rei, 2015).

Within this context, this paper tries to solve the previous limitations by presenting a new tool, ProDis, which performs prosodic acoustic dialectometric analysis easily and accurately.

1.3 Novelty

As has been already stated in Section 1.2, there is no program to date that can perform acoustic dialectometric analysis in an objective, reliable, easy and transparent way. The tool that we present solves this problem by offering the user the following possibilities.

Firstly, ProDis is objective insofar as the data it uses to perform the analysis is acoustic (i.e. the numeric values of F0, intensity and duration). Numeric data used by ProDis are evidently objective, whereas some of the previous prosodic dialectometrical studies relied on subjective intonational transcriptions.

Secondly, the analysis in entirely transparent in the sense that ProDis offers all statistical information (like distance matrices and indicators), whereas some of the previous tools did not give the user the possibility to get such pieces of information. A crucial difference in this sense between ProDis and its antecessors is that ProDis offers such statistical indicators for all the visual results (dendrograms, MDS plots, etc.) in a way that it is easy for the user to check whether the results are robust.

The third advantage of ProDis is its reliability, in the sense that its results have been tested for coherence. In order to do so, the results obtained with this software have been validated by means of perceptual data.

Last but not least, the software it is user-friendly, and fully-automatic, it has an intuitive interface, it allows to export and post-process the data and it does not require a pipeline of scripts and applications.

2 Method

The approach followed in this paper is composed of six stages summarized in the scheme illustrated in Figure 1. (A) The database consist of a set of F0 contours obtained from several speakers acquired in different cities and collected within the Catalan section of the AMPER project (Martínez Celdrán and Fernández Planas, 2003-2016). (B) The software computes the Pearson correlation between each pair of contours. (C) Given that each sentence has been recorded three times, the correlation of each sentence is computed nine times (including each possible permutation) and an average sentence correlation (ASC) score for each sentence is calculated. (D) The ASC is then evaluated for each pair of speakers of the data-base, leading to a Speaker Matrix (SM). (E) Speakers belonging to the same locale are analyzed jointly in order to get a Locale Matrix (LM), which indicates prosodic similarities between two points of survey of the AMPER database. (F) The results are finally visualized using dendrograms and multidimensional scaling (MDS) plots (in Figure 1 a dendrogram can be seen as an illustration).

The phases of the analysis pipeline are detailed in the following subsections. Namely, Section 2.1 provides a description of the database (A); Section 2.2 focusses on computation of sentences correlation (B); Section 2.3 describes ASC computation (C); Section 2.4 illustrates how the similarities matrices are created (D); Section 2.5 deals with the creation of the groups (F) and their visualization (F).

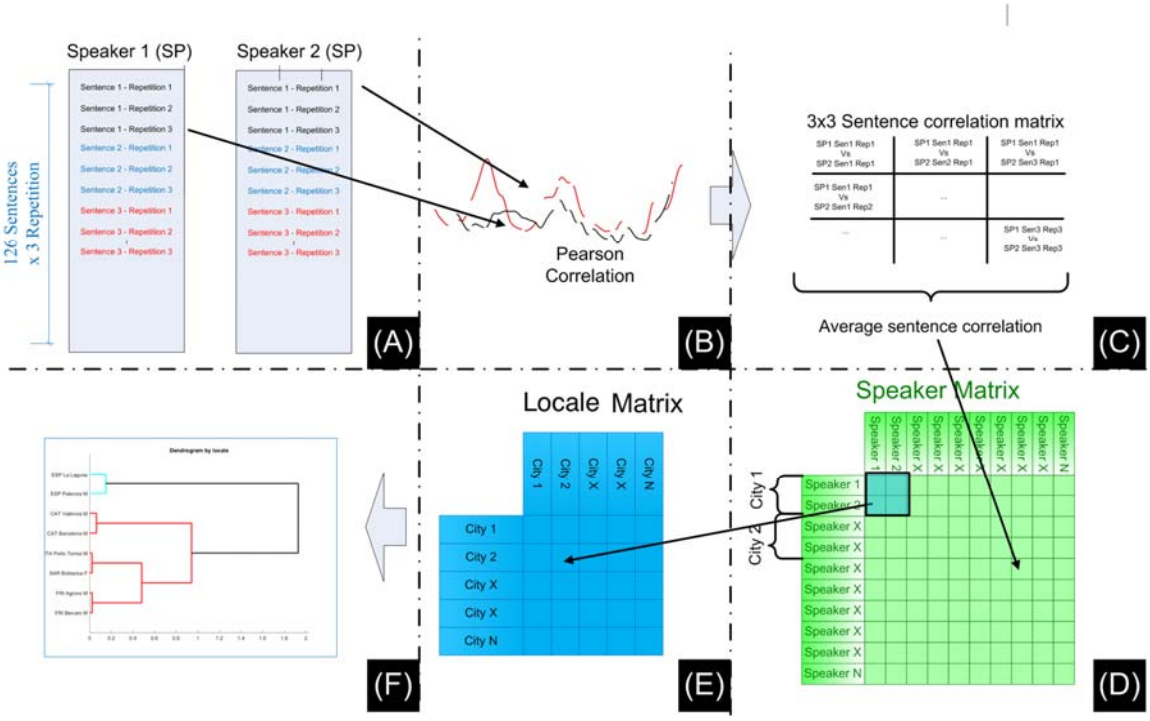

Figure 1 General schema depicting the six stages for dialectometric assessment.

2.1 Materials

In this paper two data-sets are used: one acoustic and one perceptual (described in 2.2.1 and 2.2.2, respectively). The first is used for the numerical test of the software on romance sentences, the second is used for validation purposes only.

2.1.1 Dataset-A: AMPER acoustic data

ProDis has been designed to analyze acoustic data contained in the most extended prosodic atlas available to date: the *Atlas Multimédia Prosodique de l'Espace Roman*, known as AMPER (Contini and Romano, 2002), an ongoing international project with several hundred researchers that, when it will be finished, will include data from all Romance varieties worldwide. The project is divided into subprojects, roughly speaking one for each Romance language or, for major Romance languages, one for dialectal area. In this paper we use data from different subprojects, namely AMPER-CAT (Catalan, Peninsular Spanish and some Hispano-American locales, (Martínez Celdrán and Fernández Planas, 2003-2017)), AMPER-FRIÛL (Friulian and regional Italian (Roseano and Fernández Planas, 2009)), AMPER-CAN (Canary Islands Spanish (Dorta, 2015)), AMPER-PERU (Peruvian Spanish), and AMPER-GAL (Galician; (Fernández Rei, 2011)), AMPER-AND-OR (Southeastern Peninsular Spanish

(Pàmies and Martínez Celdrán, 2015)), AMPER-ARAG (Eastern Peninsular Spanish (Simón Casas and Fernández Planas, 2015)), AMPER-MAD (Central Peninsular Spanish (Martínez Celdran, 2010)), AMPER-URU (Uruguayan Spanish (Martínez Celdrán et al., 2015)), and AMPER-EUSK (Northeastern Peninsular Spanish (Pagola Petrirena, 2002)). In addition to those locales, data from Zaragoza Spanish, Mexican Spanish, Brazilian Portuguese, Sardinian and Sardinian Italian have also been collected in order to make the results more comprehensive.

In AMPER there are usually two informants for each locale: one male and one female. Each of them reads three times a set of sentences of two different sentence-types: broad focus statements and information-seeking yes-no questions. The sentences in the corpus have a basic Subject-Verb-Object (*SVO*) structure, where both the subject and the object may be formed either only by a name (e.g. in Northern Catalan *El capità portava la caputxa* 'The captain brought the hood') or a name and an adjective/complement (Northern Catalan *El capità fenici portava la caputxa* 'The Phoenician captain brought the hood' and *El capità portava la caputxa petita* 'The captain brought the small hood'). The *V* has always the stress on the penultimate syllable, whereas the other elements of the sentence (S, O, and their adjectives/complements) have the stress position in as many places as the language allows it. In most Romance varieties, this means three positions (last syllable, penultimate syllable, and the syllable before the penultimate). Some other varieties only have two positions (e.g. last syllable and penultimate syllable, like in Northern Catalan) or only one (i.e. last syllable, like in Northern French). If a variety has three possible accentual positions, the number of sentences recorded for informant is 378 (63 sentences x 2 sentence-types x 3 recordings), which is the standard number of sentences. Nevertheless, in some locales the number of sentences is different due to the characteristics of the local variety (e.g. in the cases of varieties with less than three stress positions there will be less than 378 sentences, in the case of varieties with different syntactic interrogative structures there will more than 378; for more details about the differences in corpora see (Fernández Planas, 2005; Romano et al., 2005). Table 1 provides details about the languages and dialects, locales, number of informants and number of sentences used in this paper. On the whole, the corpus contains 36,216 sentences which constitute Dataset-A.

Language	Dialectal variety	Locales	N. Speakers	N. of sentence-types	N. of sentences
Catalan	Alguerese	L'Alguer	2	3	567
	Central	Barcelona, Girona, Tarragona, Andorra	8	3	567
	Northwestern	Lleida, Tortosa	4	3	567
	Balearic	Maó, Palma, Sant Josep de sa Talaia	6	3	567
	Valencian	València, Castelló de la Plana, Novelda	6	2	378

	Transition variety	Fraga	2	3	567
	Northern	Perpinyà	2	3	180
Spanish	Bilingual Catalan-Spanish areas	Lleida, Barcelona, Tortosa, Alacant, Palma	8	3	567
		València	2	2	378
	Central and Northern Peninsular	Logroño, Madrid, Palencia, Salamanca, Segovia, Soria, Valladolid, Zamora, Zaragoza	12	2	378
	Southern Peninsular	Almería, Bullas, Caravaca de la Cruz, Córdoba, Granada, Jaén, Málaga	13	2	378
	Canarian	La Gomera	1	2	378
	Peruvian	Lima, Trujillo, Tarapoto	3	2	378
	Mexican	México DF	2	2	378
	Uruguayan	Montevideo	2	2	378
Galician	Western	O Grove	1	2	54
	Eastern	Ribadeo	1	2	54
Portuguese	Brasilian	Pelotas	2	2	54
Friulian	Eastern	Beivars, Gradisca	4	2	378
	Northern	Agrons	2	2	378
Italian	Regional Italian (Sardinia)	Villanova Monteleone, Porto Torres	2	4	54
	Regional Italian (Toscana)	Siena, Firenze, Arezzo, Perugia	4	2	54
	Regional Italian (Friul)	Tolmezzo	1	2	54
Sardinian	Logudorese	Biddanoa Monteleone	2	4	54

Table 1 Dataset issue from AMPER project

The sentences described above are analyzed manually with a software called AMPER06 (López Bobo et al., 2007), which extracts the values of three parameters (F0, duration and intensity) for each syllable of the sentence. Due to the fact that some consonants in the sentences may be unvoiced, the software focusses on vowels.

For each vowel, AMPER06 stores a duration value (in milliseconds), an intensity value (in decibels) that corresponds to the mean intensity of the vowel, and three $F0$ values (measured at the starting point, at the midpoint and at the endpoint of the vowel). Following (t' Hart et al., 1990), $F0$ is measured in semitones instead of Hertz, which allows comparing voices of different vocal range (e.g. male and female voices). This is possible because semitones are not an absolute scale, but a relative scale. The formula used, in fact, computes semitones taking as reference value the mean $F0$ in each sentence. In other words, a given semitone value express the difference between the $F0$ of a given vowel and the mean $F0$ in the utterance.

2.1.2 Dataset-B: Comparison between acoustic and perceptual data

In order to validate the results, the data obtained with ProDis was compared with the perceived linguistic distance among the same varieties. Specifically, one of the perceptual tests described in (Fernández-Planas et al., 2017) serves to this purpose. In that test the judges listened to the intonation of two sentences and had to answer the question "Do sentence A and B belong to the same dialect?" The possible answers to the test were "yes" or "no". The stimuli were yes-no questions from the 15 speakers mentioned in Table 2 (i.e., the same speakers as in Dataset-B). A low pass filter was applied to all the stimuli in order to leave only the prosodic data. The stimuli were heard twice and distractors were included, giving as a result 121 trials that were presented with an inter-stimulus time of 800ms and a break after 25 trials.

The participants were 31 freshmen university students both male and female, bilingual of Catalan and Spanish. The results on one of the tests in Fernández Planas et al. 2017 showed that the judges were perfectly able to differentiate between their own variety and other varieties. They were also able to judge the similarity between stimuli belonging to varieties that were foreign to them (e.g. Friulian or Sardinian), but had more trouble (i.e. longer reaction times) when trying to do so.

In order to compare the results obtained by Fernández Planas et al. 2017 with the results obtained by ProDis, we selected the subset of the acoustic data presented in Table 1 that matched Fernández Planas et al. 2017 perceptual data. This reduced dataset of both acoustic and perceptual data (henceforth Dataset-B) includes nine points of survey. Specifically, Dataset-B includes y/n questions from 5 languages, 8 locales and 14 speakers (see Table 2).

Language	Dialect	Locales	N speakers
Catalan	Central	Barcelona	2
	Valencian	Valencia	2
Spanish	Canarian	La Laguna	2
	Northen Peninsular	Palencia	1
Italian	Regional Italian	Portu Turre	1
Sardinian	Logudorese	Biddanoa Monteleone	2
Friulian	Eastern	Gradisca, Beivars	4

Table 2 Number of speakers and dialectal classification in Dataset-B

2.2 Computation of sentences correlation

The acoustic difference between the prosody of two sentences can be obtained by computing the Pearson correlation, which measures the similarity between two $F0$ contours. After analyzing the AMPER data with Amper06, an $F0$ contour is converted to a vector of numeric values, i.e. of $F0$ values in semitones measured at certain points (three for each vowel). Figure 2 contains a graphic representation of two $F0$ contours of the yes-no question *¿La guitarra se toca con paciencia?* as uttered by a speaker of Spanish from Palencia (blue line) and by a speaker of Spanish of Madrid (red line). In this figure, every dot represents a $F0$ value extracted by Amper06.

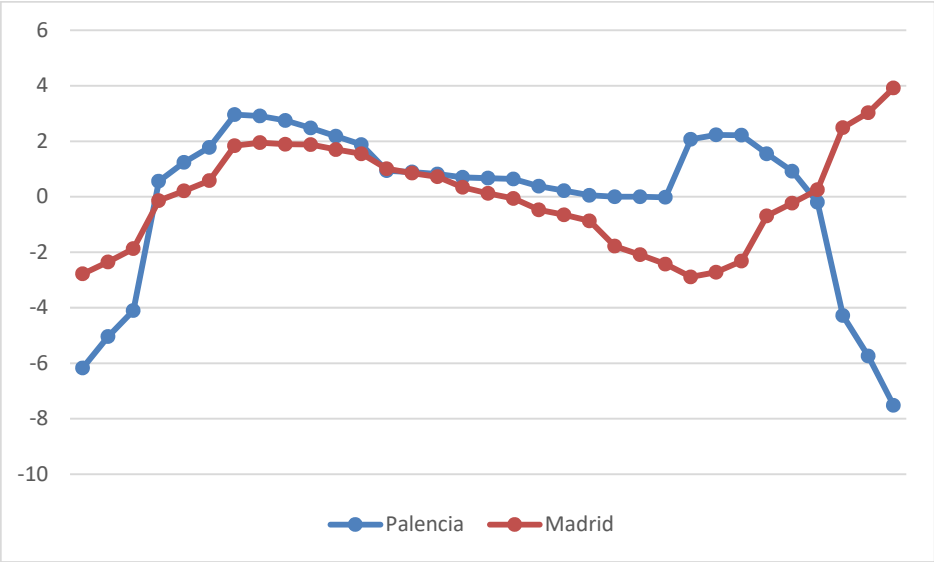

Figure 2 F0 contours of two AMPER sentences. In the x axis sample index and in the y axis F0 in semitones relative to the average F0 of each speaker.

Once the $F0$ values have been extracted, the Pearson correlation (Romano et al., 2011) is computed by applying the formula in (1).

$$R_{f_1,f_2} = \frac{\sum_i w_I(i) w_D(i)(f_1(i)-m_1)(f_2(i)-m_2)}{\sqrt{\sum_i w_I(i) w_D(i)(f_1(i)-m_1)^2 \sum_i w_I(i) w_D(i)(f_2(i)-m_2)^2}} \quad (1)$$

where $f_1(i)$ is the value of $F0$ in semitones at the curve index i; m_1 and m_2 are the average values of the contours computed in the whole sentence and the w_I is the weight due to the intensity of the signal computed as the average of the two values of energy measured in a given point expressed in decibels and w_D is the weight due to the duration of the vowel. The correlation can be weighted according to different criteria, i) duration of the vowel, ii) intensity of the vowel intensity, and iii) both of them. The correlation results of this paper are computed weighting by both duration and intensity.

2.3 ASC computation

In order to improve the intra-speaker coherence of the corpus, in the AMPER corpus, each sentence is uttered by the speaker three times, with the aim of averaging the possible variability of the speaker pronunciation. Hence, the average sentence correlation (ASC) score of each sentence is computed by considering all the possible permutation of the three sentences of the two speakers (for a total of nine permutations), and averaging their correlation results, r_{f_1,f_2} (as can be seen in Figure 1-C).

2.4 Similarities matrices creation

The ASC is computed for each pair of speakers of the data-base, leading to a Speaker Matrix (SM). This matrix represents the similarity between two speakers with the speakers grouped by similarity (Wise, 1997) (See Figure 1-D).

Then, speakers belonging to the same locale are jointly analyzed to get a Locale Matrix (LM), which indicates prosodic similarities between two points of survey of the AMPER atlas (See Figure 1-E).

2.5 Creation of the groups and visualizations

The numeric results summarized in the Locale Matrices were studied by computing cluster analysis and visualized using dendrograms and multidimensional scaling (MDS) plots.

The optimal number of clusters is determined by silhouette values (Rousseeuw, 1987). The program computes the silhouette value for the possible number of groups, then chooses the optimal value (i.e. the highest silhouette value). Such optimal number of clusters is used to visualize the groups both in the dendrogram and the MDS graph.

3 Results

This section details the results of the ProDis software applied to the AMPER atlas. The coherence of the computations is demonstrated using the perceived distances.

3.1 Acoustic prosodic similarities in Romance languages

The correlation results of dataset-A are illustrated in Figure 3. The figure displays the average distance between pitch contours acquired in different locales, computed as explained in Section 2.

The correlations among locales that appear in the map can be classified in three categories: 1) groups that display high similarities (white), 2) groups that display negative correlation values (black), and 3) groups with low correlation (gray).

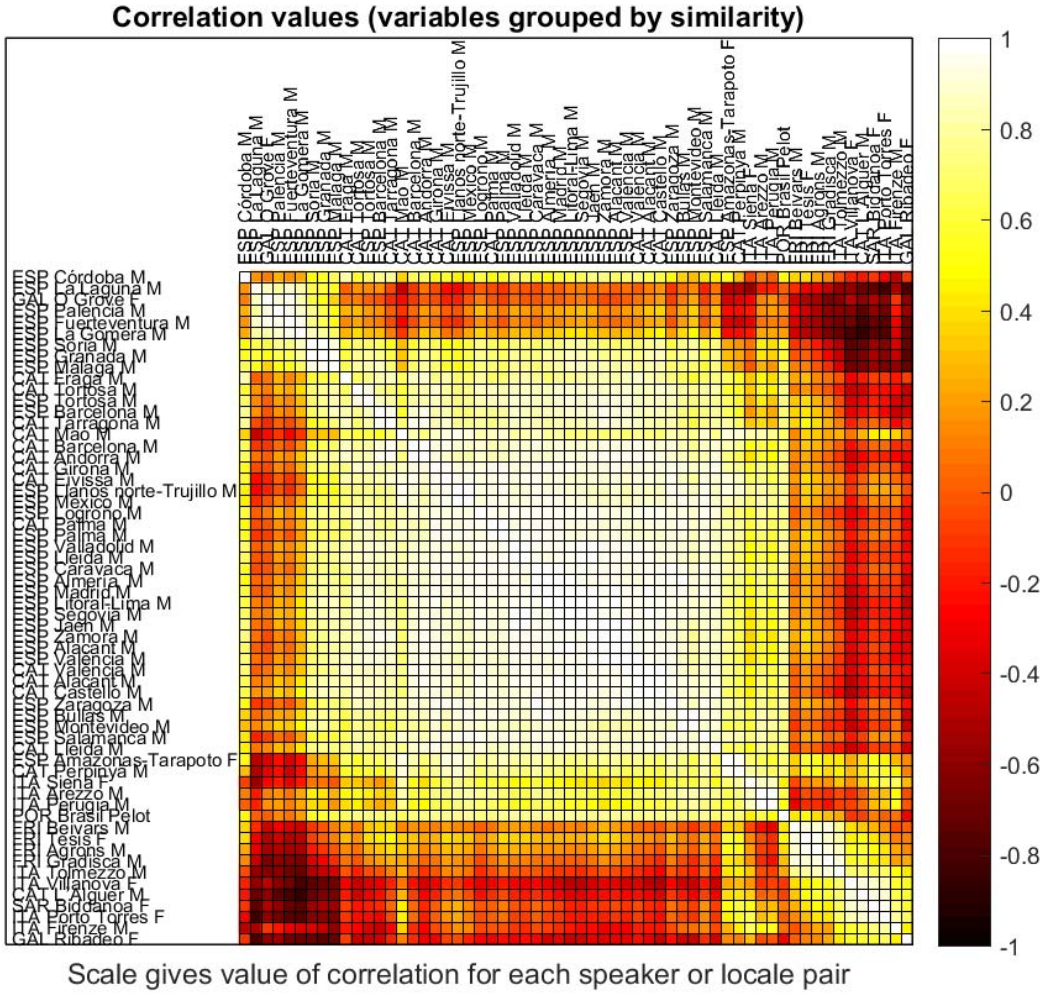

Figure 3 Correlation values between each pair of locales visualized with a color scale ranging from -1 for total dissimilarity to +1 to total similarity.

From a linguistic point of view, only groups displaying high similarities (white in Figure 3) are interesting because they are geoprosodic clusters, i.e. the clusters of dialects with similar prosodic patterns. However, deciding how many clusters there are by looking at the correlations matrix is not a good method because, on one side, it is difficult and, on the other side, it is not objective and reliable. The best method to provide a reliable and straightforward dialectal classification is implementing a silhouette values analysis to the data resulting from cluster analysis. For this reason, after computing the silhouette values for each possible number of clusters, ProDis suggests that the best choice is to split the data

into 5 groups or dialects (Figure 4), which will appear as branches of different colors in the dendrogram.

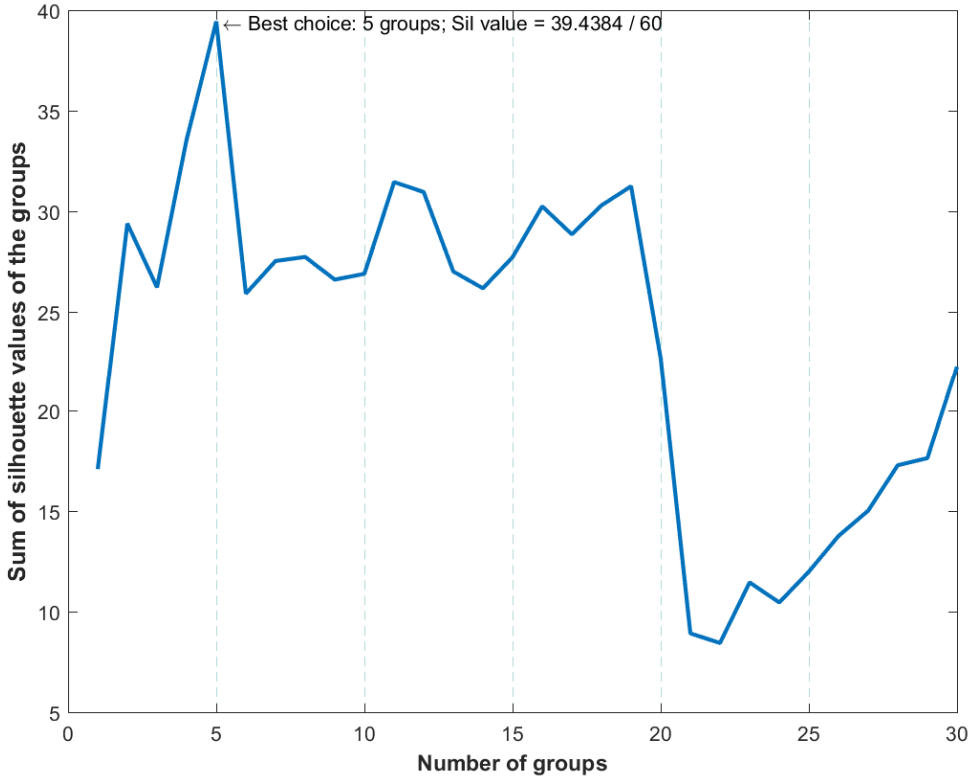

Figure 4. Silhouette scores for the cluster analysis using groups from 1 to 30

Following the silhouette assessment, the dataset-A is classified in five main branches that can be seen in Figure 5. A stylized pitch contour of a member of each group is showed in Figure 6.

1) The first group (yellow) includes several locales with falling y/n questions. Specifically, those contours are characterized by a high (or rising) last stressed syllable followed by a falling boundary tone (see Figure 6 yellow F0 contour). The group includes most Galician dialects, as well as Canary Islands and Palencia Spanish varieties (see Dorta (2013) for a description of Canarian Spanish; Fernández Rei (2011) for Galician and Martínez Celdrán and Fernández Planas, (2003-2017) for Palencia) .

2) The second group (brown) corresponds to most Spanish and Catalan dialects. The group shares an acoustic feature: in y/n questions the last final syllable is low and is followed by a final rise.
3) The third branch (green) consists of two minor sub-groups (3a and 3b). The first subgroup corresponds to Tuscan-Italian varieties. Specifically, it includes Arezzo, Perugia and Siena. The second branch includes some cities from Spanish and Catalan that are in the linguistic periphery. The differences between the groups 3a and 3b can be better appreciated in Figure 6 and are detailed in MDS discussion (next section).
4) The fourth branch (light blue) corresponds to Romance varieties spoken in Friuli, both Friulian (Agrons, Tesis, Beivars and Gradisca) and the regional variety of Italian spoken in the same region (Tolmezzo). The intonational pattern of y/n questions displays two peaks in the body of the sentence, a low last syllable and a final rise.
5) Lastly, the core of the fifth branch (dark blue) includes Romance varieties spoken in Sardinia (Logudorese Sardinian, l'Alguer Catalan and regional Italian of Sardinia). These varieties show falling patterns in y/n questions like group 1) but, in contrast with group 1), they are characterized by a falling last stressed syllable. Two locales join that core group: on one hand, Ribadeo (a Galician variety), which shows falling y/n questions, a feature shared by the Sardinian varieties. On the other hand, Firenze, which is an outlier as detailed in MDS discussion (next section).

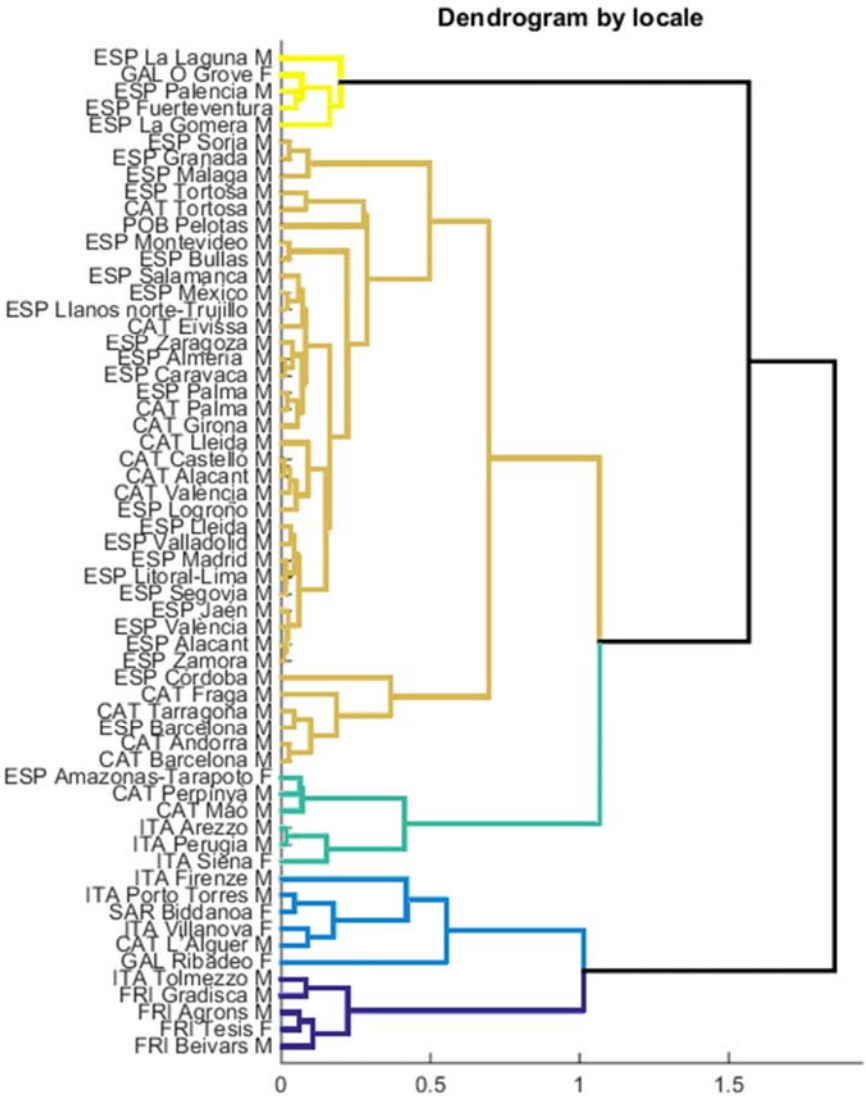

Figure 5 Dendrogram of dataset-A.

On the whole, the first linguistic conclusion one can get to is that each cluster corresponds to a specific intonational pattern. In other terms, the results offered by ProDis make sense from a linguistic point of view. Figure 6 contains an example of a pitch contour extracted from each of the main clusters of the dendrogram.

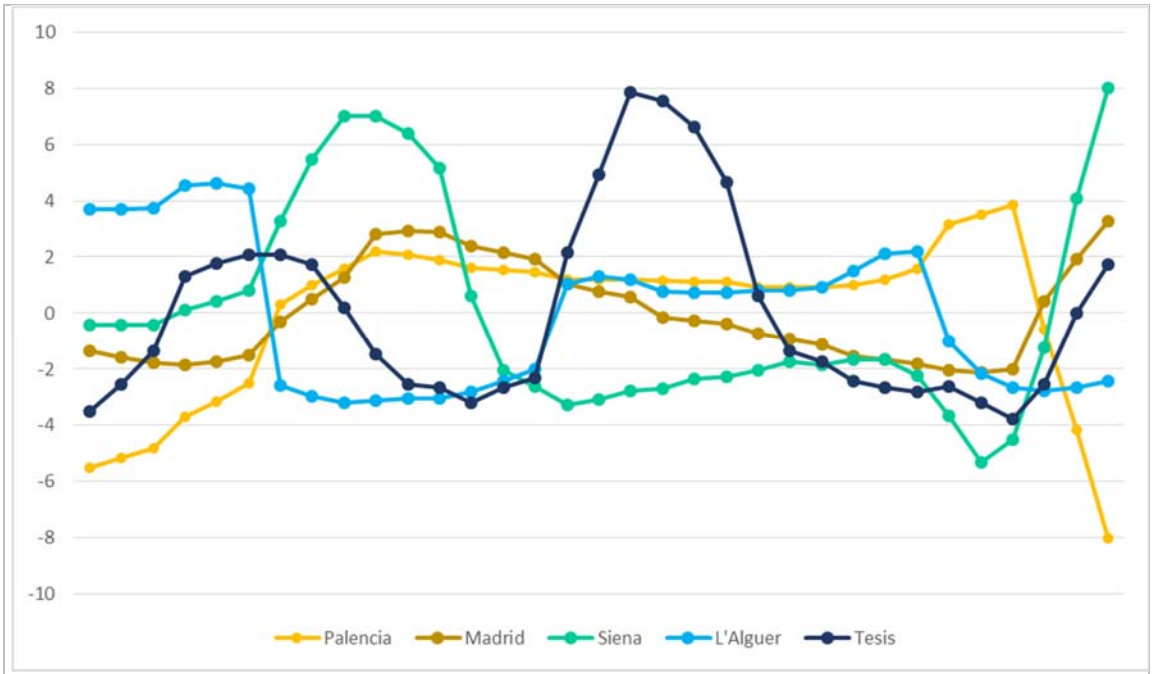

Figure 6. Syllable aligned pitch contours from a y/n question uttered by a speaker from Palencia (yellow), Madrid (brown), Siena (green), l'Alguer (light blue) and Tesis (dark blue) exemplifying the pitch pattern of five main clusters. The units in the x axis correspond to the sample index and in the y axis to the normalized frequency in semitones.

Nevertheless, from a Linguistic point of view, there are a few locales that do not match with the basic expectation of any dialectologist, i.e. that locales that are close to each other on the geographic map end up in the same cluster. In dataset-A that is the case of Firenze (Italian), Maó (Catalan), Perpinyà (Catalan) and Tarapoto (Spanish). The MDS plot turns out to be a good tool to understand the classification results of such cities.

In general, one can see that the results shown in the MDS (Figure 7) are in agreement with the dendrogram agglomerations (Figure 4). In addition, one has to point out that the reliability of the MDS analysis in question is high (stress = 0.02).

As stated above, the MDS representation allows to better understand why some locales appear as outliers in their group. For instance, Italian from Firenze was unexpectedly assigned to group 5, while a dialectologist would expect it to be assigned to group 3a because of its geographical position. The MDS plot helps to overcome this issue insofar as one can see (in upper part of Figure 6) that Firenze stands in the middle of the two clusters, and its distance with Ribadeo (group 5) and Siena (group 3a) is similar.

We have suggested that group 3 (green) can be divided into two subgroups (3a and 3b) and the MDS analysis fully supports this interpretation. It can be observed that out of the 6 points of such ensemble, Perugia, Arezzo and Siena (Italian varieties) are grouped in the upper part of the MDS, whereas the group composed of Maó (Catalan), Perpinyà (Catalan) and Tarapoto (Spanish) are in a different region of the MDS (middle part). This has a linguistic explanation: the three locales in the second group (Maó, Perpinyà and Tarapoto) have been described as peripheral dialects in their linguistic systems. Perpinyà is a Catalan variety in contact with French and therefore, with some French influence. Maó is a city of the Balearic Island and linguistically belongs to a peripheral dialect (Veny, 1993). Tarapoto is a city from the Peruvian Amazonas which Spanish is characterized by having long stressed syllables which is a shared feature with Italian data. Therefore, this group is close to the main Spanish/Catalan group but still peripheral.

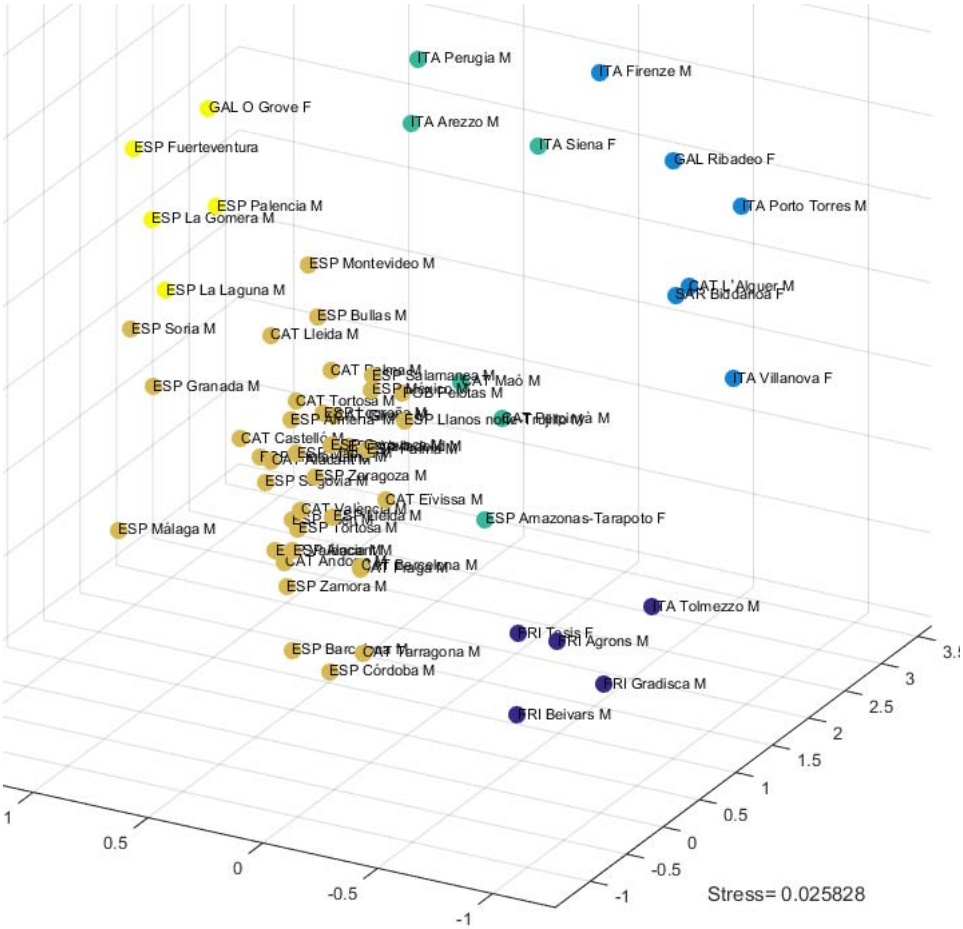

Figure 7 MDS of dataset-A. Colors correspond to the groups of the dendrogram.

3.2 Comparison to perceived distances

Following the research presented by Beijering et al. (2008), Gooskens and Heeringa (2004, 2006) and Tang and van Heuven (2007), who illustrated that dialectometric data can be used as predictor of perceived linguistic distance, in this section, the correlation automatically obtained using acoustic measurement is compared to the perceived distances. Similarly to previous research in this field, the optimal correlation between acoustics and human perception is not expected, and the comparison solely aims to illustrate the coherency and feasibility of our approach.

The acoustic and perceptual data are intrinsically inhomogeneous: perceptual data are binary information obtained answering the question "Do sentence A and B belong to the same dialect?" (see section 2.1.2). On the other hand, acoustic data are continuous variables normalized between ±1 which result from the correlation between the F0 contours of sentences A with respect to B. In order to make the results comparable (and for the purpose of this comparison only) the correlation has been discretized in binary values, assigning 1 to any correlation superior to 0.5 and inferior to -0.5 and zero otherwise. It should be noted that the city matrix resulting from the perceptual measurements contains continuous data, since each cell is the average result of several binary judgments. Hence, both acoustic and perceptual analyses are normalized between 0 and 1.

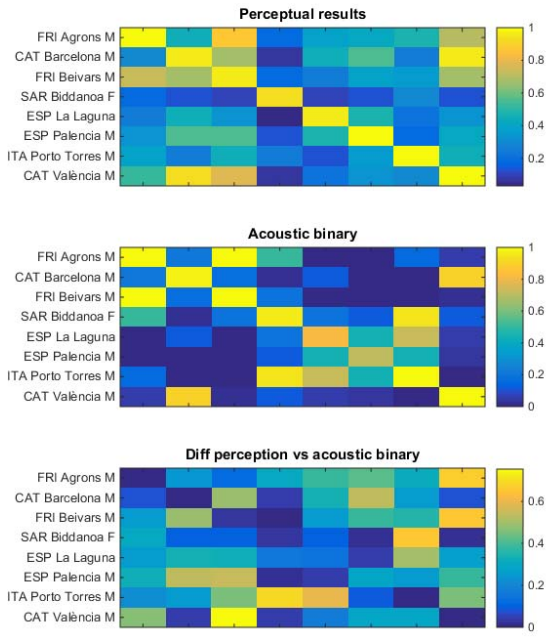

Figure 8 Locale matrices obtained from (a) perceptual and (b) acoustic data, and (c) absolute differences between (a) and (b)

Figure 8 qualitatively illustrates the city matrices of the acoustic and perceptive data and the comparison between them. Several observations can be made. see

1) First of all, the expected diagonal values of a similarity matrix is close to 1, corresponding to the fact that the pronunciation within the same city should be homogeneous. However, because of inter- and intra-speaker variation the correlation between two speakers (and even between two samples of the same speaker) is not perfect and the similarity is lower than one.
2) The variability within the diagonal of the city matrix is different in the case of acoustic and perceptual data, reaching scores of 13% (Figure 1a) and 2% (Figure 7b), respectively. This is due to the fact that human perception is categorical, i.e. listeners do not perceive gradual changes.
3) Considering 1) and 2), we expect that the difference between the acoustic and perceptive matrices (Figure 7c) would be in the best case equal to the difference between the acoustic and perceptual scores in the diagonal, hence of the order of 11%. Indeed, the average difference between them is about 27% (only 16% higher than expected value) which can be considered as moderate. In order to further evaluate the

similarity between the acoustic and perceptive matrices, a second metric is provided. The Pearson correlation score r=0.6139 (p<0.0001) is computed, and such result proves as good as the results reported for Norwegian by (Gooskens and Heeringa, 2004), which measured a correlation of r=0.67, p<.01 for lexical data. and the results reported for Chinese by Tang and van Heuven (2007) (r=.66, p<.01).

4) Most salient differences can be observed in Sardinian cities. This is because (as mentioned in section 2.1.2) Catalan listeners are not familiar with these dialects and they fail to assess their similarities and differences. In order to have a fair comparison between the computational results and the perceptive data the measurement of these cities (i.e. Sardinian from Biddanoa and regional Italian from Portu Turre) should be discarded from the comparison. Without these cities the correlation rate increases up to r= 0.8567, p<.0001.

The dendrograms of both acoustic and perceptual data are also comparable. Figure 9 shows that, when all the points are considered, the cities are grouped by pairs in the same way in both perceptual and acoustical dendrogram. However, when bigger groups are considered, the two dendograms do not fully match. If we analyze the cause it can be noted that listeners consider that Catalan and Friulian are closer than the rest of varieties and they perceived this group together with the Atlantic Spanish varieties as a group and Biddanoa and Portu Turre as separate one. Instead, acoustical results show Friulian and Sardinian varieties are considered as a group that joints Catalan points and Atlantic Spanish varieties.

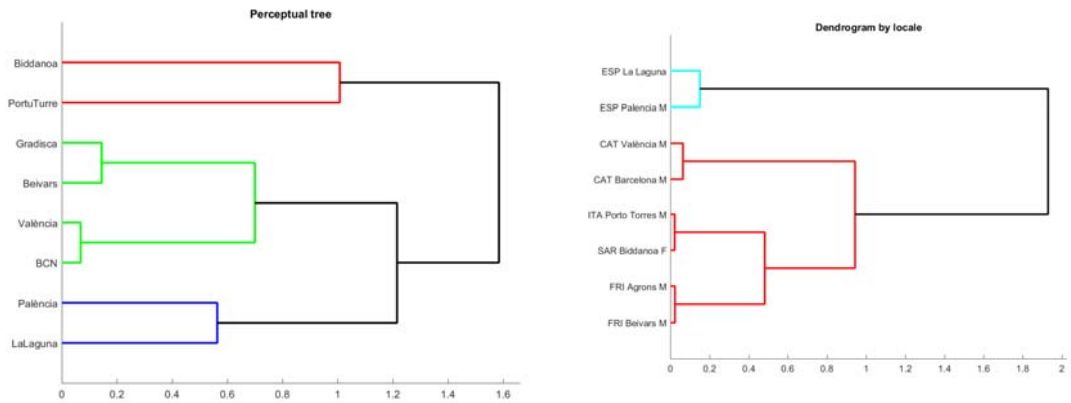

Figure 9. Dendrograms for the two types of data in Dataset-B: perceptual data (left) and acoustic data (right).

Similarly to the numeric analysis of the matrices, the results are very influenced by listeners' perception of Biddanoa and Portu Turre, which are perceived as different from the rest of varieties but also very different between them.

If the results from Sardinian are discarded, the scenario changes and not only do the groups agree in pairs but the whole dendrogram tree perfectly matches. Under this conditions, the acoustic results are perfectly mirroring human perception for known language varieties.

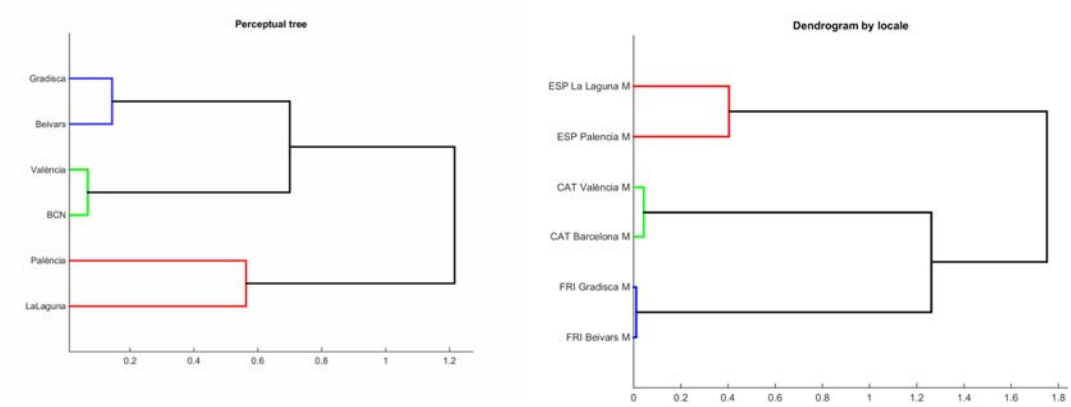

Figure 10. Dendrograms for the two types of data in Dataset-B: perceptual data (left) and acoustic data (right).

4 Conclusion

This paper has presented ProDis, a new method for dialectometry, which is used to compute prosodic distances on numeric acoustic data. ProDis computes for each point of survey in a database the similarities with all the other locales in the corpus. In order to do so, for each pair of curves the tool uses all the numeric data in the contour and performs the correlation. It is completely compatible with the AMPER database so to exploit the huge international corpus available.

The approach uses intonational values (optionally weighted by duration and intensity), therefore does not need a previous intonational transcription, it offers a complete set of transparent statistical tests, and it is user-friendly, and fully automatic and the computation does not require a concatenation of scripts and applications. Moreover, in this paper the method has proved to be reliable by correlating to the perceived distance by human listeners ($r= 0.8567$, $p<.0001$).

Therefore, as usual in dialectometry, the tool has proved to be helpful when analyzing huge datasets that cannot be dealt with by human beings. Additionally, since prosody is not usually taken into account when describing dialectal differences, the tool can make a great contribution to theoretical dialectology insofar as it can shed a new light on the phenomenon of variation.

During the past two years the tool has been under development and has been tested by several research teams. Our initial target was focused towards treating data from AMPER project, but the interest that the tool raised in several unpublished congress communications where it was presented proved that other research projects would be also interested in a dialectometric tool for prosodic acoustic data. Future developments would focus on making the software available for any kind of prosodic annotated data. This would lead to a more flexible tool.

5 Funding

This work was supported by the European Regional Development Fund (ERDF) and the Spanish Government [FFI2015-64859-P, 2016 and TIN2015-66951-C2, 2015] and the Generalitat de Catalunya [SGR 1219].

6 Acknowledgment

Dr. Josefa Dorta and Dr. Elisa Fernández-Rei for sharing the data from Canary Islands and Galicia.

7 References

Aurrekoetxea, G., Fernández-Aguirre, K., Rubio, J., Ruiz, B., Sánchez, J., 2013. “DiaTech”: A new tool for dialectology. Lit. Linguist. Comput. 28, 23–30.

Beijering, K., Gooskens, C., Heeringa, W., 2008. Predicting intelligibility and perceived linguistic distances by means of the Levenshtein algorithm. Linguist. Netherlands 25, 13–24.

Calvo, A.M., Rei, E.F., 2015. Unha ferramenta informática para a análise dialectométrica da prosodia. Estud. fonética Exp. 24, 289–303.

Contini, M., Romano, A., 2002. Atlas Multimédia Prosodique de l’Espace Roman [WWW Document]. URL http://dialecto.u-grenoble3.fr/AMPER/new.htm

Dorta, J., 2015. Atlas Multimedia de Prosodia del Espacio Románico en Canarias (AMPER-Can).

Dorta, J. (Ed.), 2013. Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela. La Página ediciones, Madrid-Sta. Cruz de Tenerife.

Fernández-Planas, A.M., Dorta, J., Roseano, P., Díaz, C., Elvira-García, W., Martín Gómez, J.A., Martínez Celdrán, E., Martínez-Celdrán, E., 2015a. Distancia y proximidad prosódica entre algunas variedades del español: Un estudio dialectométrico a partir de datos acústicos. RLA. Rev. lingüística teórica y Apl. 53, 13–45. doi:10.4067/S0718-48832015000200002

Fernández-Planas, A.M., Roseano, P., Elvira-García, W., Carrera Sabaté, J., Román Montes de Oca, D., 2017. From a perceptual point of view, is there continuity between languages in contact? Spanish Context 14.

Fernández-Planas, A.M., Roseano, P., Elvira-García, W., Cerdà-Massó, R., Romera-Barrios, L., Sabaté, J.C. i, Szmidt, D., Labraña, S., Martínez Celdrán, E., 2015b. Cap a un nou mapa dialectal del català? Consideracions a partir de dades prosòdiques tractades dialectomètricament. Estud. Fonética Exp. 24, 257–286.

Fernández Planas, A.M., 2005. Aspectos generales acerca del proyecto AMPER en España. Estud. Fonética Exp. XIV, 13–27.

Fernández Rei, E., 2011. Atlas Multimedia Prosódico do Espazo Románico (AMPER-Gal) [WWW Document]. URL http://ilg.usc.es/en/proxectos/atlas-multimedia-prosodico-do-espazo-romanico-amper-gal

Gooskens, C., 2002. The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties1 Charlotte Gooskens and Renee van. Present. Dialectol. Probl. Find.

Gooskens, C., Heeringa, W., 2006. The Relative Contribution of Pronunciational, Lexical, and Prosodic Differences to the Perceived Distances between Norwegian Dialects. Lit. Linguist. Comput. 21, 477–492. doi:10.1093/llc/fql038

Gooskens, C., Heeringa, W., 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. Lang. Var. Change 16, 189–207. doi:10.1017/S0954394504163023

Haimerl, E., 2006. Database design and technical solutions for the management, calculation, and visualization of dialect mass data. Lit. Linguist. Comput. 21, 437–444.

Heeringa, W., Johnson, K., Gooskens, C., 2009. Measuring Norwegian dialect distances using acoustic features. Speech Commun. 51, 167–183. doi:10.1016/j.specom.2008.07.006

Kessler, B., 1995. Computational dialectology in Irish Gaelic, in: Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics -. Association for Computational Linguistics, Morristown, NJ, USA, pp. 60–66. doi:10.3115/976973.976983

López Bobo, M.J., Muñiz Cachón, C., Díaz Gomez, L., Corral Blanco, N., Brezmes Alonso, D., Alvarellos Pedrero, M., 2007. Análisis y representacion de la entonación: replanteamiento metodológico en el marco del proyecto AMPER, in: La Prosodia en el Ámbito Lingüístico Románico, La Página, Santa Cruz de Tenerife, pp. 17–34.

Martínez Celdran, E., 2010. Atlas Multimedia de la Prosodia del Espacio Románico (AMPER-MAD) [WWW Document]. URL http://stel.ub.edu/labfon/amper/ampercat_resultados_cast_old.html

Martínez Celdrán, E., Coll, M., Fernández Trinidad, M., 2015. Atlas Multimedia de la Prosodia del Espacio Románico (AMPER-URUGUAY) [WWW Document]. URL http://stel.ub.edu/labfon/amper/index.html

Martínez Celdrán, E., Fernández Planas, A.M., 2003. Atlas Multimèdia de la Prosòdia de l'Espai Romànic [WWW Document]. URL http://stel.ub.edu/labfon/amper/cast/index_ampercat.html

Moutinho, L. de C., Coimbra, R., 2011. Mesure de la variation prosodique diatopique en portugais européen. Estud. fonética Exp. 20, 30–55.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., Leinonen, T., 2011. Gabmap - A Web Application for Dialectology. Dialectologia 65–89.

Pagola Petrirena, R.M., 2002. Atlas Multimedia de la Prosodia del Espacio Románico (AMPER-Eusk) [WWW Document]. URL http://amper.deusto.es/pagina/index.php?id_pagina=1

Pàmies, A., Martínez Celdrán, E., 2015. Atlas Multimedia de la Prosodia del Espacio Románico (AMPER-AND-OR) [WWW Document]. URL http://stel.ub.edu/labfon/amper/ampercat_resultados_cast_old.html

Prieto, P., Borràs-Comes, J., Roseano, P. (Coords. ., 2010. Interactive Atlas of Romance Intonation [WWW Document]. URL http://prosodia.upf.edu/iari/

Prieto, P., Cabré, T. (Eds.), 2013. L'entonació dels dialectes catalans. Publicacions de l'Abadia de Montserrat, Rubí.

Romano, A., Contini, M., Lai Jean, P., Rilliard, A., 2011. Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. Rev. Int. Lingüística Iberoam. 9.1, 13–25.

Romano, A., LAI, J.P., Roullet, S., 2005. La méthodologie AMPER. Géolinguistique 1–5.

Roseano, P., 2016. Dos décadas de dialectometría entonativa, in: A. Iglesias, A. Romero, A.E. (Ed.), Linguistic Variation in the Basque Language and Education II. Euskal Herriko UnibertsitatekoArgitalpen Zerbitzua - Servicios editoriales de la Universidad del País Vasco, Bilbao, pp. 56–80.

Roseano, P., Elvira-García, W., Fernández-Planas, A.M., n.d. Calcu-Dista: A tool for dialectometric analysis of intonational variation. Rev. Lingüística Apl.

Roseano, P., Fernández Planas, A.M., 2013. Transcripció fonètica i fonològica de l'entonació: una proposta d'etiquetatge automàtic. Estud. fonética Exp. XXII, 275–332.

Roseano, P., Fernández Planas, A.M., 2009. Atlant multimediâl de prosodie des varietâts romanichis [WWW Document].

Roseano, P., Fernández Planas, A.M., Elvira-García, W., Martínez Celdrán, E., 2015. Contacto lingüístico y transferencia prosódica bajo una perspectiva diacrónica: El caso del alguerés. Dialectol. Geolinguist. 23, 95–123.

Rousseeuw, P.J., 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Séguy, J., 1971. La relation entre la distance spatiale et la distance lexicale. Rev. Linguist. Rom. 35, 335–357.

Simón Casas, J.S., Fernández Planas, A.M., 2015. Atlas Multimedia de la Prosodia del Espacio Románico (AMPER-ARAG) [WWW Document]. URL http://stel.ub.edu/labfon/amper/ampercat_resultados_cast_old.html

t' Hart, J., Collier, R., Cohen, A., 1990. A perceptual study of intonation. Cambridge University Press, Cambridge.

Tang, C., van Heuven, V.J., 2007. Mutual intelligibility and similarity of Chinese dialects: Predicting judgments from objective measures. Linguist. Netherlands 24, 223–234.

Veny, J., 1993. Introducció a la dialectologia catalana. Enciclopèdia catalana, Barcelona.

Wise, B.M., 1997. CORRMAP Correlation map with variable grouping.