DOI: 10.26346/1120-2726-159

(Dis)Similarities between formant charts as global topological objects

Amedeo De Dominicis

Università della Tuscia, Viterbo, Italy <dedomini@unitus.it>

This paper describes a topological model to analyze the degree of similarity between formant charts. The formant charts of the stressed vowels of a number of speakers of some dialects and of a number of dialects (computed on the basis of the mean values of F_1 and F_2 of all speakers of a given dialect) represent the initial data. The model generates predictions that have been compared and validated by a perceptive discrimination test. This is a pilot study, and its major goal is methodological, rather than heuristic. Further verifications should be implemented on the same model in order to repeat and extend verification on larger corpora.*

1. Introduction

It is not unusual for speakers to express their judgments on the identity and differences between individual voices, inferring that these voices are more or less similar; analogously, they do so with linguistic varieties (e.g. dialects), claiming that these voices belong to one given dialect or another.

Traditionally, linguistics and sociolinguistics approach these perceptive judgments using a wide-ranging analysis aimed at gathering cues from several linguistic levels: phonetics, morphology, syntax, or lexicon. However, the results may be extremely complex and disjointed: the so-called isoglosses (i.e. the virtual borders that traditionally isolate one linguistic variety from another) may be dispersed and divergent, since each isogloss represents a single linguistic feature.

These perceptual judgments are most likely based on several linguistic cues, phonetic, lexical, and also syntactic, and even on some sociolinguistic stereotypes. One of these factors is the acoustic quality of the vowels, namely the relative position of the acoustic formants. This quality is traditionally mapped into formant charts (based on first and

* The Appendices of this paper are available on the journal website.

second formant values of each vowel), which represent the vocal tract, i.e. the vocal setting of the speaker(s): a geometrical representation, based on a bidimensional polygon (e.g. a trapezoid).

The formant chart can represent the voice of an individual speaker, but can also represent the vowels of a whole linguistic community/ variety, as is usual in descriptive phonetics. This paper deals with the topological analysis of these polygons. The degree of similarity between two polygons depends on the possibility of transforming the (Cartesian) space in which the polygons are inscribed: if a topological transformation is allowed, then the polygons are equivalent. Such a transformation corresponds to an equation, and in particular to polynomials of degree 2, which also provides the numerical rating of topological similarity between two polygons. This paper aims at modeling the topology of these formant polygons in order to predict the rate of similarity among the linguistic varieties that these polygons represent.

We will test this model on a few geographically adjacent Italian dialects. In order to verify their rates of topological similarity, we also conducted a perceptive discrimination test with listeners from the same dialect area. They were asked to rate the degree of similarity among the dialects. The method used to quantify the results of the perceptual test is a cluster analysis. The motivation underlying the perceptual task is to compare the prediction of the topological model with the actual judgments of the speakers. Without empirical verification, the abstract topological predictions would be mere hypotheses.

As for the topological model, the experimental sections of this paper will formulate a number of hypotheses (namely, the definitions of SI, SIms, SImd, SID, SIS, and two experimental expectations called E1 and E2, all given in §4.2) that have been tested in Appendices 1 and 2. It will be shown that the predictions of the topological model (see §4.4 and Figure 17) mainly match the results of the perceptual test (see §5.4 and Figure 18): the formant charts that are topologically more similar refer to those dialects that the listeners evaluate as more similar, and vice versa. This is even more remarkable, since we do not consider formants to be the primary acoustic correlates of differences between voices or languages.

Although the present analysis concerns some Italian dialects, the goal of this paper is methodological rather than heuristic. We focus on formant charts, because their frequency values enable us to generate a geometry, i.e. bidimensional polygons, whereas other phonetic factors cannot. The aim is to quantify the rate of topological homotopy between formant charts, that is to measure how similar the shapes of two formant charts are. We propose a new tool capable of sorting out predictions that match the impressionistic judgments of speakers asked to judge the degree of similarity between different voices or different dialects.

What are the advantages of a topological approach? Instead of comparing the acoustic profile of single vowels, as defined by their position inside two (or more) formant charts, the topological method allows one to compare – holistically – the shapes of two (or more) formant charts. The advantage comes from comparing global vocal sets rather than single points.

How is the comparison effected? Each formant chart maps a polygon representing the vowels of a speaker. We ignore consonants or suprasegmental features, and take into account only stressed vowels. This choice will be explained in §4.1. As mentioned, the formant charts may represent not only individual speakers but also linguistic varieties. In the literature, the generalization from the speakers' to the language formant chart is based on arithmetical averages and dispersion areas.[1]

In this paper, the topological model has been applied to the field of dialectometrics, on a set of geographically adjacent Italian dialects (see §4.1). The linguistic material has been chosen in order to set the most challenging test for the model: the dialects under comparison are close in both geographical and typological terms. We hope that this experiment will be repeated on larger corpora in order to validate the present model.

The paper is organized as follows. Section 2 presents the literature related to vowel system analysis, illustrates the acoustic basis of formant theory, formant chart plotting and its application to the description of linguistic variability. Section 3 presents the topological model adopted, which will be used to build a metrics relating different formant charts to one another and to calculate their rate of similarity. Section 4 provides an empirical application of the topological model to the particular field of dialectometrics. All computations and formulas concerning the application to the corpus are given in Appendices 1 and 2; however, the predictive value of our model lies precisely in the validation of these formulas and their numerical results. Section 5 describes the perceptive experiment carried out in order to establish a terminus *a quo*: that is, the judgments of the listeners regarding the degree of similarity of the selected dialects. Section 6 sums up the results of our research. Section 7 recapitulates and suggests other possible linguistic domains in which to apply the model.

2. Analysis and representation of vowel systems

In this section we present a vowel system analysis, its acoustic base, and the formant chart plotting.

2.1. Source-filter theory and speaker variability

Formant frequencies are one of the acoustic correlates of differences between vowels and are primarily determined by shape characteristics of the vocal tract. They are identified by the peaks in the spectral envelope of the speech signal and determined by the natural resonances of the vocal tract. For any given speaker, changes in formant frequencies depend primarily on changes in the shape and position of the articulators (tongue, lips, jaw, etc.), according to the so-called source-filter theory developed by Gunnar Fant (Fant 1960, 1966).

The frequency of the formants depends on the length and the diameter of the vocal tract, especially on the interaction between the two main spaces in the vocal tract. As for the two lowest formant frequencies, their frequency varies as follows. A mouth constriction lowers F_1 and raises F_2. This creates a less compact vowel spectrum. The acoustic energy is spread out over both low and high frequencies, as in the vowels [i] and [e]. A pharyngeal constriction raises F_1 and lowers F_2. This makes the vowel spectrum more compact, as in the case of [ɑ] or [o]. Actually, the source-filter mathematical formalism is more complex: the relation between the formant frequencies and the dimension of vocal areas is interactive rather than merely linear; however, a linear approximation is a reasonable one for a topological processing of data stemming from fairly close dialects and homogenous classes of speakers.

2.2. Formant chart

A formant chart (Potter & Steinberg 1950) is a schematic plotting of the vowels' formant frequencies resulting in different vowels. The vertical axis of the diagram denotes vowel height, with high vowels at the top of the diagram, whereas the horizontal axis indicates the anterior/posterior space, with front vowels to the left of the diagram. The vertical axis represents the values (in Hz) of the first formant F_1 (in reverse order, i.e. with lower values corresponding to high vowels and vice versa) while the horizontal axis represents the values of the second formant F_2 (with higher values corresponding to front vowels and vice versa). The first two formants are important in determining the quality of vowels and are frequently said to correspond to the open/closed and front/back dimensions.

Although F_2 is somewhat related to the degree of vowels backness, the distance between F_2 and F_1 (i.e. the subtraction F_2-F_1) is a better predictor (e.g. Fant 1973, Lindau 1978, Goldstein 1984, Rosner & Pickering 1994). Moreover, there are numerous strategies for improving formant plots. Ladefoged (1967: 87) used the Mel scale and other scholars have used the Bark scale or the ERB scale.

Our aim is to plot a homogeneous formant representation for any voice/dialect. What we need is a consistent measurement strategy. The association of front/back vowels to the values of F_2 is sufficient for our purpose, although probably not the best predictor. Our aim is to compare two or more formant polygons using a topological equation. Irrespective of plotting F_1 *vs* F_2 or F_1 *vs* F_2-F_1, the equation (i.e. the topological transformation from one formant polygon to another) yields the same result in both cases: plotting the values of F_1 and F_2 in Mel or in Bark (rather than in Hz) or plotting the value of F_2-F_1 (instead of F_2) homotopically translates, but does not change the overall shapes of both polygons. We provide a detailed comparison in Appendix 3.

As explained earlier, each point in a formant chart ideally corresponds to a given articulatory location. Figure 1 illustrates an example of a formant chart for four vowels and the correlation between the articulatory space and the formant frequencies as plotted in a formant chart.

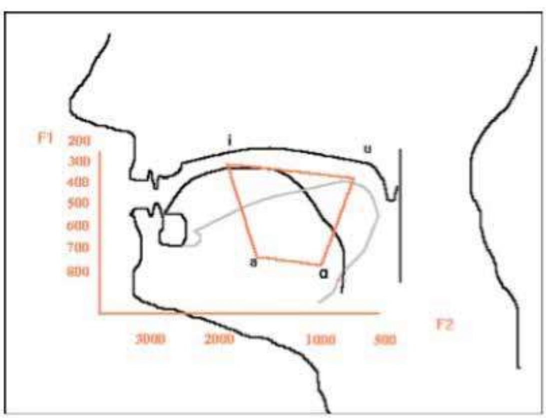

Figure 1. Correlation between the articulatory space and the formant frequencies as plotted in a formant chart in which F_2 is on the x-axis, and F_1 is on the y-axis.

The following vowel chart (Figure 2) shows the frequency regions for F_1 and F_2 relating to seven Italian vowels as pronounced by one speaker. Owing to phonetic variation, vowel location is not represented by points but by dispersion areas.

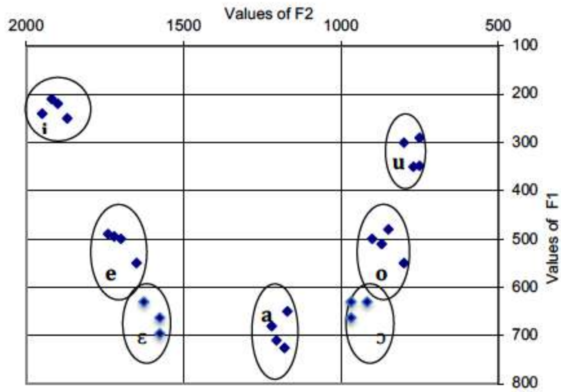

Figure 2. An example of a formant chart and of vowel dispersion areas in Italian (author's voice).

2.3. Linguistic differences in vowel production

As for the differences in language with regard to formant frequencies, an example is given in Figure 3 (adapted from Bradlow 1993) showing the differences in the /i/, /e/, /o/, and /u/ locations in Greek, Spanish, and English, and plotting them in a formant chart, i.e. the area covered by the quadrilaterals defined by the mean F_1 and F_2 values of the four common vowels.

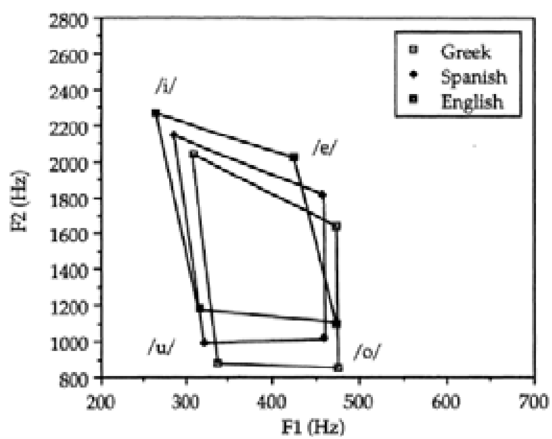

Figure 3. The /i/-/e/-/o/-/u/ areas in English, Spanish, and Greek. All tokens in all languages appear in CVCV contexts (adapted from Bradlow 1993).

Traditionally, the differences in vowel production related to different speakers or languages are accounted for by separately comparing the frequency differences relative to the individual vowels. However, this overlooks the fact that the vowel space as a whole depends on the given articulatory setting, as determined by linguistic and anatomical constraints, such as age, sex, mood, health, and speech habits (Abercrombie 1967: 7-9; Laver & Trudgill 1991: 235-264).

In order to overcome this limitation, we propose a new approach and a new tool. The crux of our approach is to holistically compare the shapes of the different vowel spaces as represented by the quadrilaterals (polygons) defined by the mean F_1 and F_2 values. This approach adopts the concept of topological deformation of the space. It enables one to measure the topological identity of two formant charts and to determine the numerical rate of similarity between them. The next section describes the method.

3. Topological approach

In this section we elaborate a model that will be used to compare different formant charts and the polygons they generate, as well as the metrics used to evaluate their similarity.[2]

3.1. Continuity-Deformation

In order to account for the differences among speakers and dialects, we adopt the concept of topological deformation. In *On Growth and Form* (Thompson 1917), the biologist D'Arcy Wentworth Thompson used this concept to describe changes in shape – morphology – of various living creatures (see Figure 4).

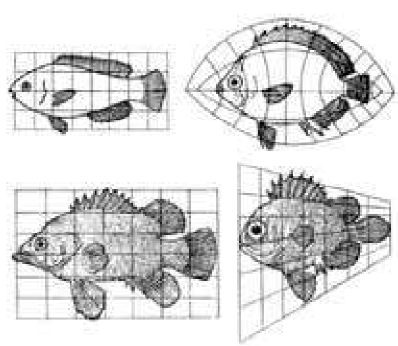

Figure 4. D'Arcy Thompson's illustration of the shapes of fish: his transformation grids aimed to show how a simple mathematical operation could turn a parrotfish (top left) into an angelfish (top right), or a Polyprion (bottom left) into a big-eye (bottom right).

Two forms are topologically equal if they retain the same function even if they are geometrically different, and if a transformation exists that leads from one to another. For instance, a cup and a ring (or a torus) are topologically equal, because both have a hole; an '8' and an eyeglasses frame are topologically the same because both have two holes. In both examples, a transformation (by means of an equation) leads from one form to the other.

John O'Connor and Edmund Robertson, mathematicians at the School of Mathematics and Statistics at the University of St. Andrews, Scotland, implemented some of these ideas in their MAC application *D'Arcy Thompson's Pictures* (DTP – <www-groups.dcs.st-and.ac.uk/~john/darcy.html>).[3] Such an application enables users to alter pictures in real time by varying parameters in mathematical functions. DTP uses quadratic maps, that is, maps of the form $f(x,y) = (p(x,y), q(x,y))$ where p and q are polynomials of degree 2 in two variables. Hence the user has the freedom to vary 10 parameters. With so many degrees of freedom, quadratic maps allow one to vary the parameters (the values of x and y) continuously and to follow the results of such variations.

In the next section (§3.2) we show the user interface of DTP by means of an example.

3.2. Implementation of DTP on a case study

DTP may, for instance, be implemented in the case of the English and Spanish vowel space (already described in Figure 3). Figure 5 shows the relative user interface. The picture in the left-hand square is mapped to the image in the right-hand square by the map defined at the bottom. The user can alter the parameters in the two functions which define the transformation by pressing the buttons below them, and thus observe the picture change continuously. The rate at which the parameters change can be varied, as can the range of the x and y values.

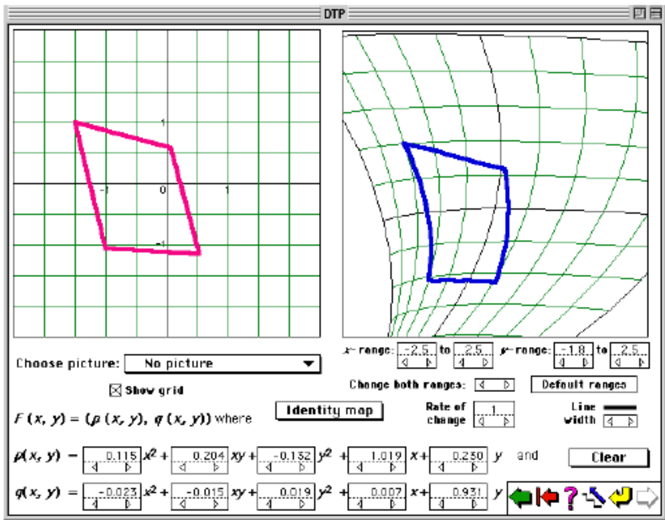

Figure 5. Implementation of DTP on the vowel space in English and Spanish.

The topological transformation shown in Figure 5 can be described as follows:

– given the Cartesian coordinates x (abscissa) and y (ordinate), and a Cartesian plane with x and y values between -2.5 and +2.5;
– assuming the values of the Cartesian axes are modified so that x ranges between -2.5 and +2.5; and y ranges between -1.8 and +2.5;
– given F(x,y) as the function that describes all objects inscribed into the given Cartesian space;
– then, the equation that describes the topological transformation of the curve in the left-hand square into the resulting one (in the right-hand square), is the following:
  $F((0.115x^2 + 0.204xy - 0.132y^2 + 1.019x + 0.230y), (-0.023x^2 - 0.015xy + 0.019y^2 + 0.007x + 0.931y))$

This formula describes not only the transformation in Figure 5, but the homotopic deformation of every object that is inscribed in the same topological space.

3.3. Provisional discussion

The equations describe the transformations and are a mathematical representation of a given articulation setting (referring to a given language, or speaker). Representing an articulatory space usually involves using electromagnetic techniques, a standard method for measuring the position of parts of the mouth, which gives a GEOMETRICAL picture. The above equations may be converted into a picture of the same articulators, but their representation is TOPOLOGICAL, not geometrical. This means that different articulatory spaces may be discovered as geometrically different but topologically identical. A typical example of this divergence between geometry and topology is the case of a torus (a kind of ring) compared to a mug: they are geometrically different but topologically identical.

3.4. Equivalence class and its establishment

The acoustic data of our corpus will be plotted on a formant chart for each speaker, along with the corresponding vowel formant space (i.e. the so-called vowel trapezoid, or – better – vowel polygon). The goal is to compare these vowel polygons (say two polygons: A and B) by means of the transformations (T) in order to establish an equivalence class. According to the topological theory, this comparison may or may not involve the identity of two polygons, depending on whether or not it is possible to obtain both a T and a reverse-T (T^{-1}) for each pair of polygons: if a polygon A can be transformed into a polygon B, and also B into A, then A and B belong to the same equivalence class. If not, then A and B do not belong to the same class. The ideal, theoretical case of equivalence class between A and B is represented by the following equations: T(A) = B, T(B) = A, and reverse-T[T(A)] = A, or reverse-T[T(B)] = B.

In our case, however, this classical path cannot be followed, for two reasons. Firstly, when we plot A into DTP, we draw a figure that looks like A, but it is not necessarily A: we obtain another figure, say A'. As a consequence, A and A' may not be exactly the same. The same goes for B and B'. Secondly, when we apply a transformation T to A' by means of DTP, and generate a figure that looks like B, then what we obtain is not exactly B, but something (hopefully) as similar as possible to B, say B''. B and B'' may not be exactly the same. The same goes for T(B') generating not exactly A, but A'' (a polygon as similar as possible to A). This potentially imperfect coincidence between A and A', A'' and A, B and B', B'' and B depends on the operating principle of DTP: since it produces continuous deformations of space, most results actually derive from the operator's choices, manual skill and manipulations.

As a consequence, in our model, an equivalence class cannot be demonstrated by means of the canonical equations, because – in light of the above – these equations should be reformulated as follows: T(A') = B'', T(B') = A'', and reverse-T[T(A')] = A'', or reverse-T[T(B')] = B''. In order to fill the (at least potential) gap between A-A'-A'' and B-B'-B'', we adopt two principles: a REJECTION STATEMENT and a SIMILARITY INDEX.

The rejection statement, or rejection clause, claims that if B'', i.e. the polygon T(A'), and the polygon B are patently different, then they are not considered as belonging to the same class. This statement avoids any further measurement between two figures that are not patently equal.

Moreover, if the rejection statement does not apply (i.e. if T does generate a figure more or less similar to what is expected), then we must obtain a measure of that similarity. This issue refers to ranking among different degrees of equivalence: the similarity or equivalence between the polygons T(A') and B may be more or less strong, depending on how much B'' and B differ, and consequently on the 'gap' that the topological transformation (T) must bridge. This gap may be more or less large, and accordingly T may be more or less heavy. A measure of T-heaviness is given by comparing the default-T or zero-T that describes the null-T or degenerated-T between A' and A' (i.e., T(A') = A'), with the actual T that describes A' and B''. Let us consider the following examples.

3.5. Examples of comparison and the definition of the Similarity Index (SI)

Figure 6 shows the formant chart of three polygons. Let us suppose that we want to transform the dash-dot-line polygon A into the dashed-line one B. In Figure 7 we find the original polygon A' on the left and the B' on the right. B' = T(A'). B' should be identical to the dashed-line polygon B in Figure 6, but actually it is not. Thus, the dash-dot-line polygon A and the dashed-line one B will not be considered as belonging to the same class.

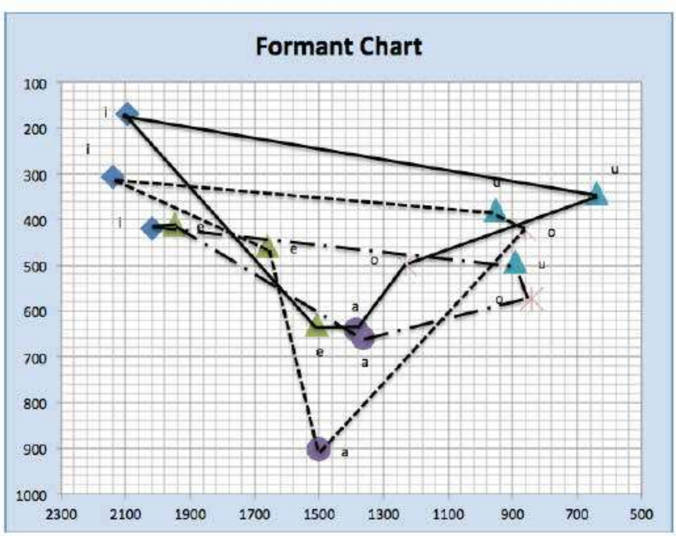

Figure 6. Formant chart of three polygons.

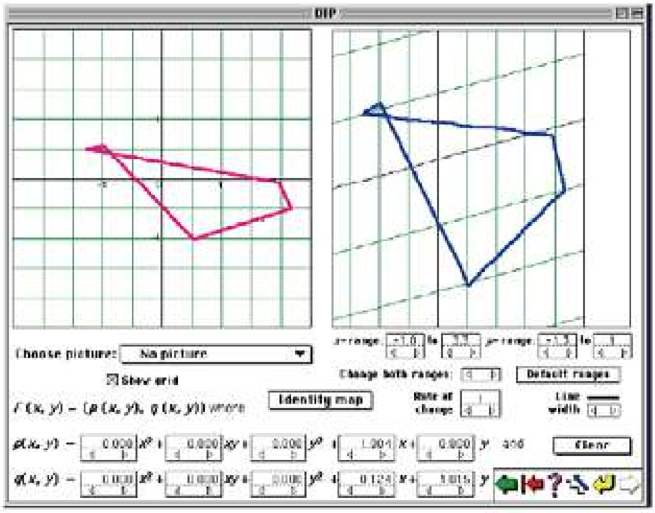

Figure 7. Implementation of DTP on two polygons presented in Figure 6: T of the dash-dot-line polygon (left) into the dashed-line one (right).

Next, let us consider the formant chart in Figure 8 and suppose that we wish to transform the dash-dot-line polygon A into the dashed-line one B.

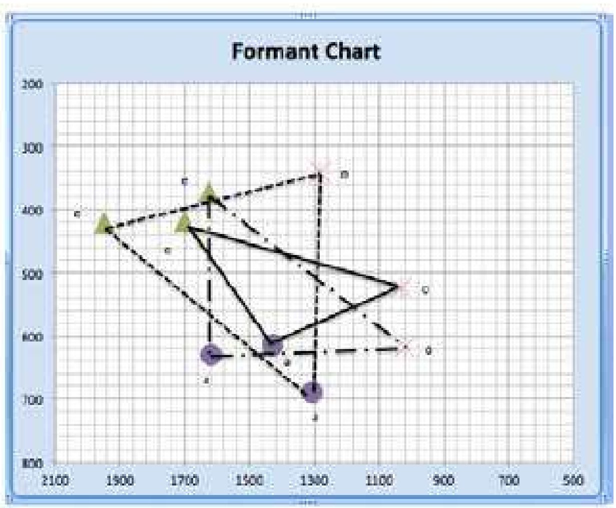

Figure 8. Formant chart of three polygons.

In Figure 9 we find the original polygon A' on the left and B'' on the right. B'' = T(A'). In Figure 10 we find the reverse-T: B' is on the left and A'' on the right. A'' = reverse-T(B').

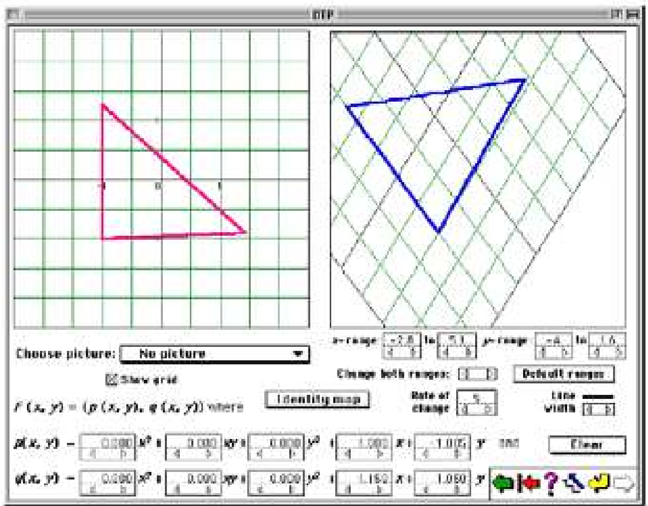

Figure 9. Implementation of DTP on two polygons presented in Figure 8: T of the dash-dot-line polygon (left) into the dashed-line one (right).

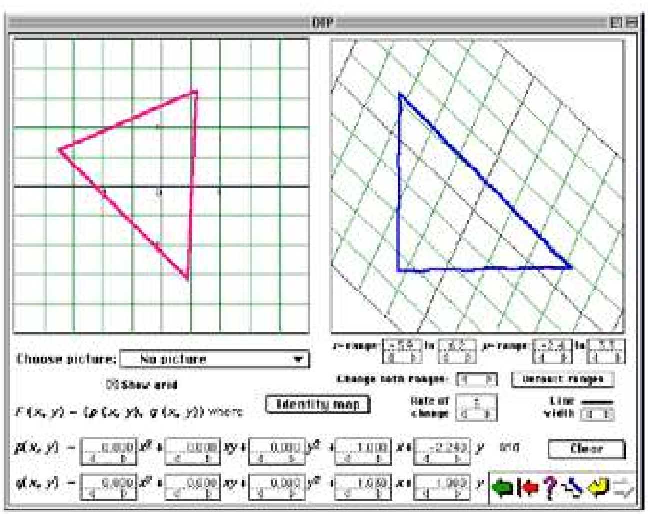

Figure 10. Implementation of DTP on two polygons presented in Figure 8: T of the dashed-line polygon (left) into the dash-dot-line one (right).

A" in Figure 10 and B" in Figure 9 are not patently different from, respectively, the dash-dot-line (A) and the dashed-line (B) polygons in Figure 8. Thus, A and B may belong to the same class. Now, however, the degree of equivalence between them, i.e. the SIMILARITY INDEX (SI), must be evaluated. To this purpose we will compare the equations that describe T in DTP: in particular, the values of x-range, y-range, and the function F that by default[4] is $((0x^2 + 0xy + 0y^2 + 1x + 0y), (0x^2 + 0xy + 0y^2 - 0x + 1y))$ are what we call 'default-T' or 'zero-T' (where the starting and final polygons are by definition the same); the values of the actual T are the second term of comparison. In the default-T all variables are 0 or 1. Thus, by comparing default-T and actual-T, we obtain a numerical measure that depends on the value of the variables of actual-T. In particular, it depends on the Δ between the value of each variable in default-T and in actual-T. The Δ refers only to NATURAL NUMBERS, i.e. numbers that represent the absolute value of the difference between variables, regardless of their algebraic sign.

In short, we obtain the following data:

Δ_1 (x-range) = Δ between the value of x-range in actual-T and in default-T $(-2.5 \leq x \leq 0)$

Δ_2 (x-range) = Δ between the value of x-range in actual-T and in default-T $(0 \leq x \leq +2.5)$

Δ_3 (y-range) = Δ between the value of y-range in actual-T and in default-T $(-2.5 \leq y \leq 0)$

Δ_4 (y-range) = Δ between the value of y-range in actual-T and in default-T $(0 \leq y \leq +2.5)$

Δ_5 = Δ between the value of each of the five variables in actual-T and in default-T $(0x^2 + 0xy + 0y^2 + 1x + 0y)$

Δ_6 = Δ between each of the five variables in actual-T and in default-T $(0x^2 + 0xy + 0y^2 + 0x + 1y)$.

The SIMILARITY INDEX (SI) is defined as follows: $SI = \Sigma (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Sigma\Delta_5 + \Sigma\Delta_6)$

The closer the SI value is to 0, the more similar the polygons. In the example in Figure 9, SI = 7.585 is calculated as follows: $(0.3+2.6+1.5+0.9+1.085+1.15+0.05)$, where

$\Delta_1 = 0.3$

$\Delta_2 = 2.6$

$\Delta_3 = 1.5$

$\Delta_4 = 0.9$

$\Sigma\Delta_5 = (0x^2 + 0xy + 0y^2 + 1x + 1.085y) = 1.085$

$\Sigma\Delta_6 = (0x^2 + 0xy + 0y^2 + 1.150x + 1.050y) = 1.15+0.05$

In the example in Figure 10, SI = 11.27, calculated as follows: (3.4 + 3.7 + 0.1 + 0.8 + 2.24 + 1.03), where:

$\Delta_1 = 3.4$
$\Delta_2 = 3.7$
$\Delta_3 = 0.1$
$\Delta_4 = 0.8$
$\Sigma\Delta_5 = (0x^2 + 0xy + 0y^2 + 1x + 2.240) = 2.24$
$\Sigma\Delta_6 = (0x^2 + 0xy + 0y^2 + 1.030x + 1y) = 1.03$

As a consequence, the polygons in Figure 9 are more similar than the polygons in Figure 10, and the equivalence class in Figure 9 is stronger than the one in Figure 10.

3.6. Method

In order to assign the SIs of T to assess the identity of two polygons (and the vowels they represent), an appropriate method must be elaborated.

The transformations on polygons representing formant spaces are very special figures. In fact, differently from other figures submitted to a topological transformation, they have no ‘holes’: thus, their transformations are not subordinated to the restriction concerning the invariance of the number of holes between the original and the transformed figure.

Moreover, they are special because inversion by symmetry is not allowed. For phonetic reasons, this transformation is illegitimate in the case of polygons representing formant spaces: a [u] will always be on the right of [i], and an [a] always below [u] and [i].

Given A as the polygon representing the formant space of the first voice and B as the polygon representing the formant space of the second voice, then T_1 is the transformation from A to B, and T_2 is the transformation from B to A.

Next, provided there is no rejection, due to the implementation of our rejection statement, T_1 is defined by its SI (SI_1) and thus T_2 by its SI_2. The experimental expectation is that, if the values of SI_1 and SI_2 are very similar (and close to 0), then A and B are similar and compatible.

As for the degree of similarity among dialects (not speakers), we calculated the arithmetic mean of the Hz values of the formants of all the speakers of the same dialect, then plotted the corresponding mean formant chart for each dialect and compared the corresponding polygons (each polygon representing a given dialect) by means of a topological transformation. For instance, given a dialect x and a dialect y, we compare the forms of the respective polygons representing their formant charts and calculate the SI of the transformations leading from x to y (SI_1) and from y to x (SI_2). If the values of these SI are close to 0 and similar, then we state that x and y are similar. In order to assess the significance of this conclusion, we compared the mean SI of the transformation among individual speakers belonging to a given dialect x (SIS) and the mean SI of the transformations between dialect x and dialect y (e.g. the mean between SI_1 and SI_2). We symbolize the former as SIS and the latter as SID. The experimental expectation is that, if the SIS value is greater than the SID value, then the result is not significant and not validated (because the degree of similarity among speakers of the same dialect is lower than among dialects); on the contrary, if the SIS value is smaller than the SID value, then the result is significant and the degree of dialectal similarity is validated (because the degree of similarity among speakers of the same dialect is higher than the one among dialects).

In §4 we will apply the model to a dialect corpus and assign a numerical rating of similarity among the given dialects.

In §4.2 we will integrate the formal details of the model for dialectometric purposes: the index of affinity between the formant charts of two speakers or of two languages/dialects will be called SIms (referring to a pair of speakers) and SImd (referring to a pair of dialects); SID (referring to a given dialect) is the arithmetic mean among the values of all SImd concerning all pairs of dialects in which a given dialect occurs, and SIS is the arithmetic mean among the values of all SIms concerning all speakers of a given dialect.

Additionally, in §4.2 we will implement the mathematical relationships among these indexes and their value into two experimental expectations (E1 and E2).

These definitions and expectations are the formal framework of the topological model. In addition, we will provide an empirical validation of the model, by comparing our results to the speakers' perceptive judgment by means of a discrimination test (§5).

4. Empirical application to a dialect corpus

In §3.6 we described the general model, providing a metrics relating different formant charts and calculating their degree of similarity. This model can be implemented in a number of empirical applications. In §7 we propose a provisional list of possible purposes. In order to adapt the model to a given field, we must adopt some additional definitions. This section provides an empirical application of the model to the particular field of dialectometrics.

4.1. Data

The object of this analysis is the stressed vowels of four dialects from southern Lazio and Molise: the dialects of Alvito (Frosinone), Montaquila (Isernia), Veroli (Frosinone), and Formia (Latina). They are geographically close to one another and located across the border between central and upper-southern Italian dialects.

The speech corpora of these dialects are stored in the 'Centro di documentazione sui dialetti del Lazio meridionale' (University of Cassino). The corresponding audio files (digitized at a 44100 Hz sampling rate) have been made available thanks to the center administrators and colleagues Luca Lorenzetti, Paolo Milizia, and Giancarlo Schirru.

In order to enable an adequate comparison among speakers voices, only female voices were selected, since comparing male and female voices would involve major anatomical differences. The speakers list is the following:

Alvito: EL (age 55), PAS (age 63), ROS (age 58)
Montaquila: LOR (age 54), SER (age 23), MAR (age 26), ID (age 66), GI (age 53)
Veroli: AZ (age 69), AS (age 63)
Formia: AN (age 55), VIN (age 57)

All speakers completed only primary school. No further anagraphic or socio-cultural information on the subjects is available in the archive of the 'Centro di documentazione sui dialetti del Lazio meridionale'.[5] Hence, we only note the speakers' initials, education level, and age. The number of speakers per location varies depending on the available recordings.

In order to facilitate the analysis of the acoustic formants, only stressed vowels have been analyzed. However, the consonant context of each vowel is roughly comparable in all dialects (the consonant before each given vowel is the same, or homorganic; if possible, the same applies to the consonant following the vowel).

Two main reasons explain the limitation to stressed vowels. First, from an acoustic standpoint, measuring formants is not easy in the case of unstressed vowels, since these bear less energy and the formants may be invisible on a spectrogram; also, one might mistake for a formant a resonance that is slightly stronger than the others, simply because such resonance is a multiple of a lower harmonic of an unpredictable nature. Second, unstressed vowels are sometimes realized as fricative consonants, i.e. with no periodicity whatsoever. Moreover, unstressed vowels are often neutralized and centralized, so that the formant chart of unstressed vowels risks reducing, if not to a single point, at least to a a small polygon.[6]

4.2. Definitions and experimental expectations

The empirical application to dialectometrics requires prior implementation of the following seven definitions (D1-D7):

D1. Polygon-S (or Speaker Polygon): this is the formant chart of the mean F_1 and F_2 values of one speaker. If three speakers are analyzed (e.g. p1, p2, p3) for a given dialect, then we have three Polygon-S and six T (see below), thus six SI.

D2. Polygon-D (or Dialect polygon): this is the formant chart of the mean F_1 and F_2 values of all speakers of one dialect. If three speakers are analyzed for a given dialect and we obtain three Polygon-S, we also obtain one Polygon-D for that dialect.

D3. T (and T^{-1}): T is the topological transformation from a polygon A to a polygon B. T^{-1} (or reverse-T) is the opposite, that is, the T from B to A. This pair of Ts applies both to Polygon-S and to Polygon-D. For each T, an SI (Similarity Index) is calculated. Thus, for each pair of T and T^{-1} we obtain two SI. The arithmetic mean of the values of these two SI is an index of affinity between two speakers (for Polygon-S) or two dialects (for Polygon-D). This arithmetic mean is called SIm (SImd or SIms; see below).

D4. SImd (between a pair of dialects): this is the arithmetic mean between the values of two SI (stemming from a T and T^{-1} pair) concerning two Polygon-D. If two dialects (and two Polygon-D) are analyzed, we obtain two SI (SI_1 and SI_2) but only one SImd, calculated by the formula $(SI_1 + SI_2)/2$.

D5. SID (of a given dialect): this is the arithmetic mean among the values of all SImd concerning all pairs of dialects in which a given dialect occurs. If there are 4 dialects, each dialect entails three SImd; thus the SID of each dialect is calculated using the formula $(SImd_1 + SImd_2 + SImd_3)/3$.

D6. SIms (between a pair of speakers): this is the arithmetic mean between the values of two SI (stemming from a T and T^{-1} pair) concerning two Polygon-S belonging to the same dialect. For instance, if in a given dialect two speakers are analyzed and we obtain two Polygon-S, then we obtain two SI (SI_1 and SI_2) but just one SIms, calculated using the formula $(SI_1 + SI_2)/2$.

D7. SIS: this is the arithmetic mean among the values of all SIms concerning all speakers of a given dialect. For instance, if in a given dialect three speakers are analyzed (e.g. p1, p2, p3), then we obtain three pairs of T/T^{-1} and three pairs of SI, from which we obtain three SIms; the SIS of that dialect is calculated using the formula $(SIms_1 + SIms_2 + SIms_3)/3$.

The analysis is based on two experimental expectations (E1 and E2):

E1. Comparing the SImd of two dialects, the closer to 0 is the SImd value, the more similar the two dialects are.

E2. SIS < SID: that is, the experimental expectation E1 must be previously verified by the E2. E2 verifies that the degree of similarity among the speakers of a given dialect is higher than the degree of similarity among the dialects, the dialects' degree of similarity should be lower than the degree of similarity among the speakers of each dialect. The verification of E2 ensures the significance of E1.

4.3. Formant charts

The formants were measured close to the center of the vowel using Praat. Praat has an inbuilt formant tracker, based on LPC, but it often makes mistakes. Such errors were checked and corrected, when necessary, by visual inspection of the so-called 'Spectral Slices' – based on FFT (Fast Fourier Transformation).

We did not take into account low-mid stressed vowels (ɛ, ɔ). There is no phonological opposition between high-mid and low-mid Italian vowels in most Italian dialects. Aside from this empirical consideration, the topological comparison of formant charts concerns shapes or figures, not points. This means that comparison 'by shape' allows one to background the number of the vertices of the formant polygon: this avoids false diagnoses of dissimilarity (based on a high SI, but merely due to comparing vocalic systems with a different number of vowels). In other words, even if dialect x only has a mid vowel and no high-mid *vs* low-mid vowels, all these vowels are virtually contained 'inside' the formant space (or polygon).

The same considerations apply to more complex vowel systems (with phonological central vowels and phonological rounded *vs* unrounded vowels, both front and back). However, should the internal vowels (i.e. front rounded and back unrounded vowels) form a kind of 'hole' inside the polygon of the cardinal or corner vowels, then the vowel polygon turns out to be a toroid (a kind of ring). In this instance, topological comparison is possible only with another vowel polygon with a hole, because no topological transformation, by definition, exists between a figure with a hole and a figure without holes. This restriction is a basic postulate of topology (see §3.6).

4.3.1. Dialect of Alvito

Three speakers were analyzed in Alvito: EL, PAS, ROS. The values of the formants are given in Tables 1-4. In all tables, σ stands for Standard Deviation.

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
trite ('crushed')	4	i	294 (3.03)	2630 (6.72)
deva ('he gave')	3	e	412 (3.09)	2227 (21.92)
fame ('hunger')	5	a	783 (11.14)	1522 (29.42)
criatura ('child')	4	u	335 (3.84)	1030 (8.07)
avema ('we have')	4	ɛ	685 (3.56)	1916 (10.82)
tosta ('hard')	3	ɔ	463 (4.49)	1054 (4.18)
sciote ('unpackaged')	3	o	417 (4.02)	1054 (4.32)

Table 1. Formants F_1 and F_2 (in Hz) of stressed vowels (underscored) by PAS (Alvito).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
risa ('laughs')	6	i	274 (7.60)	2641 (13.04)
abbedeva ('he lived')	3	e	438 (10.20)	2663 (6.12)
levate ('taken off')	3	a	702 (6.12)	1662 (6.54)
revenuta ('come back')	3	u	233 (8.99)	478 (5.71)
severe ('harsh')	4	ɛ	600 (30.82)	2071 (6.86)
madonna ('madonna')	5	ɔ	662 (7.20)	1213 (9.22)
chemenione ('holy communion')	4	o	314 (3.56)	1111 (7.25)

Table 2. Formants F_1 and F_2 (in Hz) of stressed vowels (underscored) by EL (Alvito).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
rite ('rituals')	3	i	396 (3.68)	2745 (26.77)
sendeva ('he heard')	3	e	478 (4.08)	2030 (8.16)
incontravamo ('we met')	4	a	586 (5.21)	1484 (5.53)
menuto ('minute')	5	u	396 (11.24)	1111 (15.01)
festa ('holiday')	5	ɛ	560 (14.14)	1928 (11.80)
Antonio ('Antonio')	2	ɔ	560 (10)	1011 (9)
stagione ('summer')	3	o	457 (7.76)	1082 (8.99)

Table 3. Formants F_1 and F_2 (in Hz) of stressed vowels (underscored) by ROS (Alvito).

Vowels	F1 (σ)	F2 (σ)
i	321 (53.42)	2672 (51.81)
e	442 (27.14)	2306 (264.48)
a	690 (80.84)	1556 (76.54)
u	321 (67.24)	873 (281.25)
ɛ	615 (52.12)	1971 (70.40)
ɔ	561 (81.24)	1092 (86.88)
o	396 (60.23)	1082 (23.27)

Table 4. Mean values of the formants F_1 and F_2 (in Hz) of stressed vowels in Alvito (by EL, PAS, and ROS).

Variance and σ in Table 4 are both below the vowel formant perceptual discrimination threshold (Liu & Kewley-Port 2004, Oglesbee & Kewley-Port 2009). Figure 11 shows the formant charts of all speakers of Alvito.

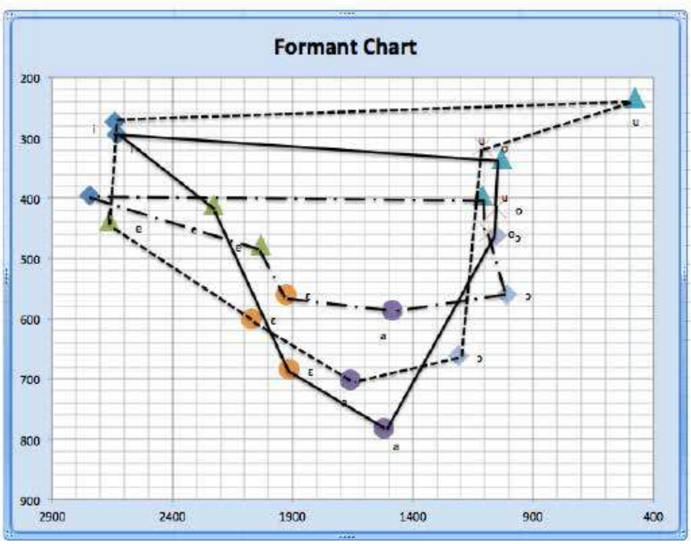

Figure 11. Formant chart of Alvito: EL (dashed-line), PAS (solid line), ROS (dash-dot line).

4.3.2. Dialect of Montaquila

Five speakers were analyzed in Montaquila: LOR, SER, MAR, ID, GI. The values of the formants are given in Tables 5-10.

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
marite ('husband')	5	i	217 (10.24)	2557 (8.06)
devene ('they gave')	3	e	438 (8.60)	1941 (8.57)
stavamme ('we stayed')	3	a	600 (8.16)	1054 (12.28)
mmenuta ('she came')	3	u	241 (7.78)	808 (9.39)
festa ('holiday')	4	ɛ	437 (10.20)	1948 (7.69)
dopo ('after')	4	ɔ	498 (11.18)	852 (7.87)
maggior ('greater')	3	o	389 (4.10)	1131 (9.41)

Table 5. Formants F_1 and F_2 (in Hz) of stressed vowels by LOR (Montaquila).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
marite ('husband')	4	i	437 (10.91)	2405 (9.73)
devi ('you must')	3	e	498 (8.73)	2253 (9.39)
stavamme ('we stayed')	4	a	797 (8.63)	1357 (12.07)
venuta ('she came')	3	u	417 (10.61)	1128 (10.27)
primavera ('springtime')	5	ɛ	596 (14.40)	1940 (14.14)
dope ('after')	4	ɔ	663 (6.41)	1164 (12.21)
giorno ('day')	5	o	424 (12.12)	1121 (15.01)

Table 6. Formants F_1 and F_2 (in Hz) of stressed vowels by SER (Montaquila).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
stritte ('narrow')	4	i	394 (13.97)	2276 (17.80)
deva ('he gave')	3	e	437 (8.16)	2208 (8.99)
javamme ('we went')	4	a	715 (11.29)	1575 (14.56)
menuta ('she came')	5	u	368 (14.91)	1207 (18.28)
traversene ('they cross')	2	ɛ	468 (10)	2231 (11)
donna ('woman')	5	ɔ	640 (14.14)	1223 (19.55)
juorne ('day')	5	o	422 (16.88)	1228 (14.26)

Table 7. Formants F_1 and F_2 (in Hz) of stressed vowels by ID (Montaquila).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
marite (‘husband’)	4	i	237 (18.45)	2035 (8.33)
devo (‘I must’)	3	e	372 (10.70)	1671 (8.98)
eravamo (‘we were’)	4	a	685 (10.55)	1581 (16.44)
perdute (‘lost’)	3	u	215 (14.71)	618 (19.29)
vera (‘true’)	5	ɛ	528 (21.42)	1536 (21.09)
Andonja (‘Antonia’)	3	ɔ	416 (16.53)	1178 (16)
mezzogiorno (‘midday’)	4	o	372 (14.79)	1021 (14.30)

Table 8. Formants F_1 and F_2 (in Hz) of stressed vowels by GI (Montaquila).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
iscritta (‘enrolled’)	5	i	381 (15.92)	2611 (15.86)
vedeva (‘he saw’)	3	e	507 (13.58)	1984 (18)
faciavamme (‘we made’)	4	a	752 (23.37)	1380 (15.65)
menute (‘they came’)	4	u	368 (21.77)	1198 (15.15)
diverse (‘different’)	3	ɛ	573 (12.02)	2066 (16.53)
songe (‘I am’)	5	ɔ	573 (33.97)	1290 (19.70)
giorni (‘days’)	5	o	414 (21.15)	1220 (20)

Table 9. Formants F_1 and F_2 (in Hz) of stressed vowels by MAR (Montaquila).

Vowels	F1 (σ)	F2 (σ)
i	333 (88.89)	2376 (207.41)
e	450 (48.88)	2011 (209.09)
a	709 (66.45)	1389 (192.24)
u	321 (79.07)	991 (237.01)
ɛ	520 (60.40)	1944 (229.67)
ɔ	558 (91.36)	1141 (151.22)
o	404 (20.36)	1144 (75.70)

Table 10. Mean values of the formants F_1 and F_2 (in Hz) of stressed vowels in Montaquila (by LOR, SER, MAR, ID, and GI).

Variance and σ in Table 10 are both below the vowel formant perceptual discrimination threshold (Liu & Kewley-Port 2004, Oglesbee & Kewley-Port 2009). Figures 12-13 show the formant charts of all speakers of Montaquila.

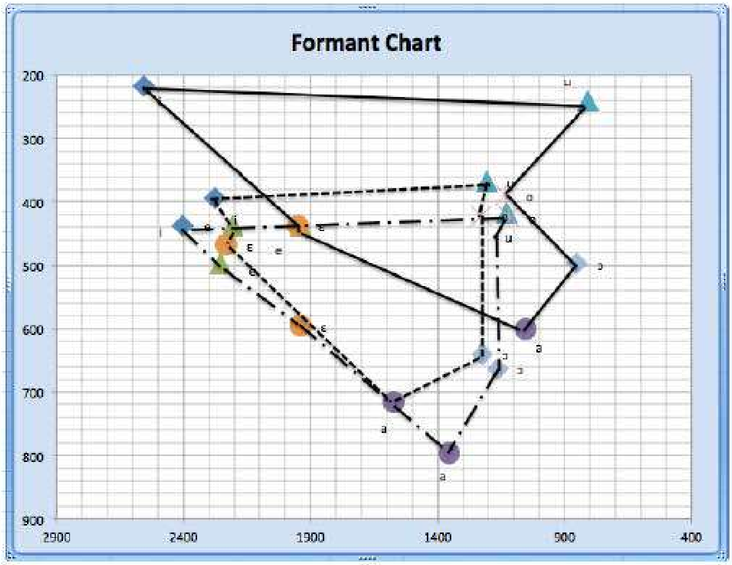

Figure 12. Formant chart of Montaquila: LOR (solid line), SER (dash-dot line), ID (dashed-line).

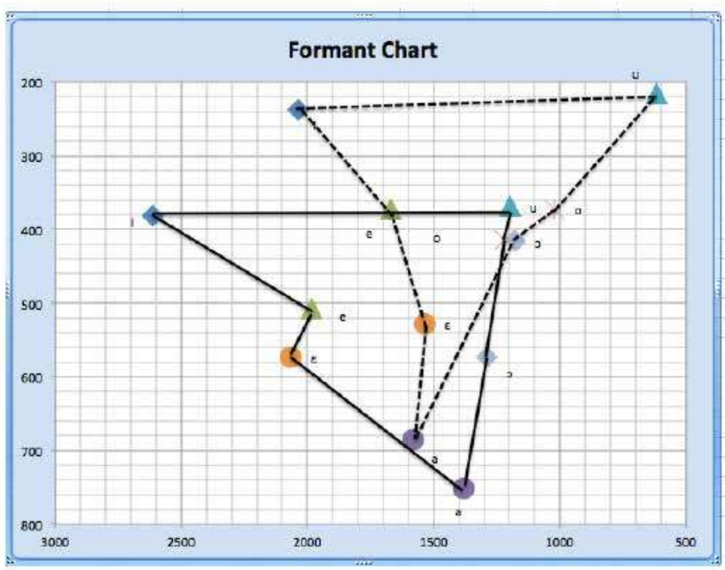

Figure 13. Formant chart of Montaquila: MAR (solid line), GI (dashed-line).

4.3.3. Dialect of Veroli

In Veroli, two speakers were analyzed: AS, AZ. The values of the formants are given in Tables 11-13.

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
marito ('husband')	5	i	322 (16.88)	2467 (138.83)
vedevo ('I saw')	3	e	404 (8.16)	2614 (10.70)
stavamo ('we stayed')	4	a	726 (13.66)	1103 (67.68)
tenuto ('held')	2	u	275 (5)	565 (15)
America ('America')	3	ɛ	475 (16.08)	2231 (126.87)
nonna ('grandmother')	5	ɔ	601 (138.57)	1107 (131.35)
giorno ('day')	5	o	432 (16.41)	1133 (106.37)

Table 11. Formants F_1 and F_2 (in Hz) of stressed vowels by AS (Veroli).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
marito ('husband')	4	i	225 (11.51)	2955 (132.94)
benedetta ('blessed')	4	e	371 (78.13)	2565 (76.97)
eravamo ('we were')	3	a	761 (102.19)	1590 (89.81)
Rufa (first name)	5	u	421 (142.37)	1130 (131.30)
festa ('holiday')	4	ɛ	496 (136.16)	2467 (117.11)
madonna ('madonna')	3	ɔ	473 (12.02)	1265 (88.41)
giorno ('day')	5	o	420 (18.09)	1297 (123.73)

Table 12. Formants F_1 and F_2 (in Hz) of stressed vowels by AZ (Veroli).

Vowels	F1 (σ)	F2 (σ)
i	273 (48.5)	2711 (244)
e	387 (16.5)	2589 (24.5)
a	743 (17.5)	1346 (243.5)
u	348 (73)	847 (282.5)
ɛ	485 (10.5)	2349 (118)
ɔ	537 (64)	1186 (79)
o	426 (6)	1215 (82)

Table 13. Mean values of the formants F_1 and F_2 (in Hz) of stressed vowels in Veroli (by AS, and AZ).

Variance and σ in Table 13 are both below the vowel formant perceptual discrimination threshold (Liu & Kewley-Port 2004, Oglesbee & Kewley-Port 2009). Figure 14 shows the formant charts of all speakers of Veroli.

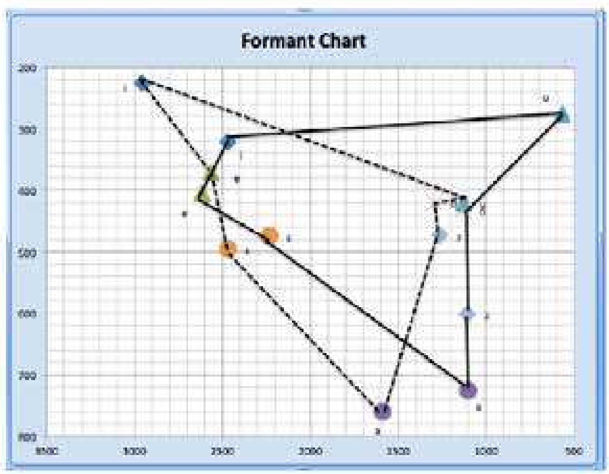

Figure 14. Formant chart of Veroli: AS (solid line), AZ (dashed-line).

Figure 15 shows the formant charts of the dialects of Veroli, Montaquila, and Alvito (Tables 4, 10, 13). The formant chart of Formia is shown in Figure 16.

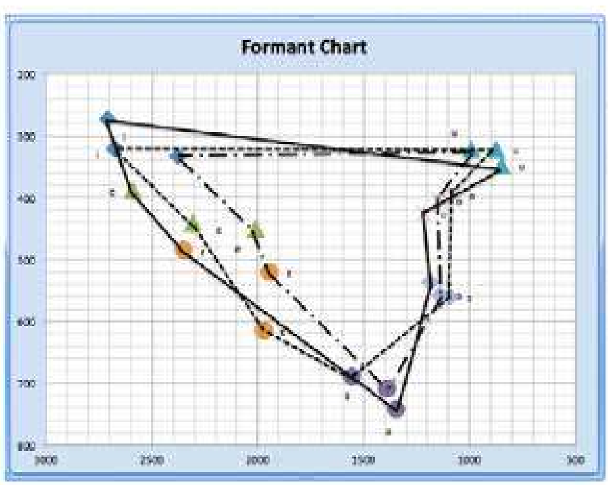

Figure 15. Formant chart: mean values of Veroli (solid line), Alvito (dashed-line), Montaquila (dash-dot line). The formant chart of Formia is shown in Figure 16.

4.3.4. Dialect of Formia

In Formia, two speakers were analyzed: AN, VIN. The values of the formants are given in Tables 14-16.

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
mariteme ('my husband')	4	i	315 (11.51)	2436 (108.24)
steve ('I stayed')	5	e	507 (143.44)	2141 (155.64)
vale ('it is worth')	3	a	874 (84.56)	1528 (69.08)
venuta ('come back')	3	u	456 (54.46)	780 (61.64)
araperta ('open')	2	ɛ	670 (60)	2182 (58)
nonna ('grandmother')	5	ɔ	670 (157.45)	977 (135.33)
stagione ('summer')	5	o	498 (146.02)	948 (148.61)

Table 14. Formants F_1 and F_2 (in Hz) of stressed vowels by AN (Formia).

Words	Occurrences	Vowels	F1 (σ)	F2 (σ)
mariti ('husbands')	5	i	303 (24.96)	2590 (141.43)
steve ('I stayed')	3	e	466 (83.05)	2141 (68.24)
eravamo ('we were')	4	a	1060 (117.47)	1446 (114.40)
venuto ('come back')	4	u	303 (111.73)	794 (112.77)
naperto ('he did not open')	3	ɛ	548 (82.05)	1790 (88.75)
nome ('name')	5	ɔ	670 (141.14)	1146 (144.30)
giovane ('young')	4	o	507 (82.14)	1038 (114.16)

Table 15. Formants F_1 and F_2 (in Hz) of stressed vowels by VIN (Formia).

Vowels	F1 (σ)	F2 (σ)
i	309 (6)	2513 (77)
e	486 (20.5)	2141 (0)
a	967 (93)	1487 (41)
u	379 (76.5)	787 (7)
ɛ	609 (61)	1986 (196)
ɔ	670 (0)	1061 (84.5)
o	502 (4.5)	993 (45)

Table 16. Mean values of the formants F_1 and F_2 (in Hz) of stressed vowels in Formia (by AN, and VIN).

Variance and σ in Table 16 are both below the vowel formant perceptual discrimination threshold (Liu & Kewley-Port 2004, Oglesbee & Kewley-Port 2009). Figure 16 shows the formant charts of all speakers of Formia, and the relative formant chart of the dialect of Formia (solid line).

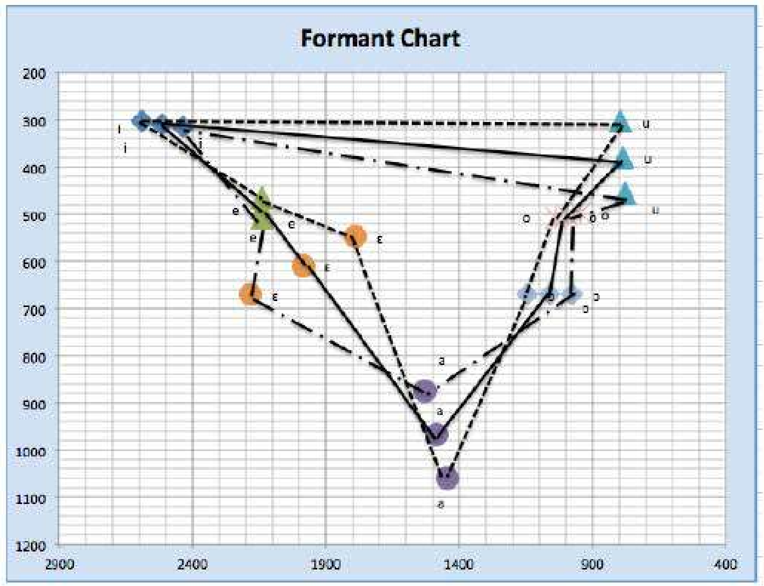

Figure 16. Formant chart of Formia: AN (dash-dot line), VIN (dashed-line), Mean values (solid line).

4.4. Equations and computations

In Appendices 1 and 2 we calculate the SI of the topological transformations relating the formant charts of speakers and dialects. Appendix 1 shows comparisons among speakers of the same dialect. Appendix 2 shows comparisons among dialects and takes into account the mean formant charts (i.e. the formant charts of the mean values of the stressed vowels of a given dialect). In Appendix 1 we list the topological transformations, the related equations, and the SI concerning the Polygon-S of the speakers of Montaquila, Alvito, Veroli, and Formia. In Appendix 2 we list the topological transformations, the related equations, and the SI concerning the Polygon-D of the dialects of Montaquila, Alvito, Veroli, and Formia.

All T presented in Appendices 1 and 2 passed the rejection statement. As expected (according to E1), the values of SImd are smaller for Alvito-Montaquila (0.807) and larger for Veroli-Alvito (1.412) and Veroli-Montaquila (2.551); they are small also for Formia-Alvito (1.070) and Formia-Montaquila (0.5785), whereas SImd is larger for Formia-Veroli (1.6855). This means that the dialects of Formia, Alvito and Montaquila are topologically predicted as being more similar than that of Veroli: the dialects of Alvito and Montaquila and the dialects of Formia and Montaquila are topologically closer, whereas the dialect of Veroli is topologically more distant. Figure 17 shows a graph reporting SImd values for each pair of dialects:

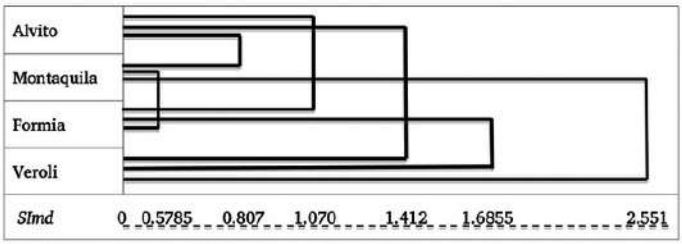

Figure 17. Graph derived from the topological distances showing the clustering of dialects.

These results confirm the linguistic behavior of the listeners in the perceptive discrimination tests (to be presented in §5). A divide distinguishes the dialect of Veroli from the dialect cluster of Alvito-Montaquila (these two dialects appear very similar, and both differ from the dialect of Veroli). Moreover, the dialect of Formia is very similar to the dialect cluster of Alvito-Montaquila, but differs from the dialect of Veroli.

The computation of SImd enables the calculation of the SID of all dialects:

SID Formia: $(1.070 + 1.6855 + 0.5785)/3 = 1.111$
SID Alvito: $(0.807 + 1.412 + 1.070)/3 = 1.096$
SID Veroli: $(2.551 + 1.412 + 1.6855)/3 = 1.882$
SID Montaquila: $(2.551 + 0.807 + 0.5785)/3 = 1.312$

The computation of SIms (and SIS) stems from the SI of the transformations between the Polygon-S of the speakers, given above. According to the experimental expectation E2, if SIS < SID, then the results of the previous analysis of SImd will be validated. This is indeed the case, as shown below. The values of SIms and SIS of Formia are the following:

T (AN → VIN): SI $= (0.4 + 0.2 + 0 + 0 + 0.283 + 0.209) = 1.092$
T^{-1} (VIN → AN): SI $= (0 + 0 + 0.4 + 0 + 0.154 + 0.169) = 0.723$
SIms and SIS of the speakers of Formia: 0.907 (< SID Formia = 1.111)

The values of SIms and SIS of Veroli are the following:

T (AZ → AS): SI = (0+0+0+0+0.221+0.420) = 0.641
T^{-1} (AS → AZ): SI = (0+0+0+0+0.135+0.397) = 0.532
SIms and SIS of the speakers of Veroli: 0.586 (<SID Veroli = 1.882)

The values of SIms and SIS of Alvito are the following:

T (EL → PAS): SI = (0+0+0+0+0.229+0.150) = 0.379
T^{-1} (PAS → EL): SI = (1+0+0+0+0.383+0.177) = 1.56
SIms (EL -PAS): 0.969

T (EL → ROS): SI = (0+0+0+0+0.441+0.398) = 0.839
T^{-1} (ROS → EL): SI = (0+0+0+0+1.008+0.995) = 2.003
SIms (EL -ROS): 1.421

T (PAS → ROS): SI = (0+0+0+0+0.240+0.469) = 0.709
T^{-1} (ROS → PAS): SI = (0+0+0+0+0.265+0.781) = 1.046
SIms (PAS-ROS): 0.877
SIS of the speakers of Alvito: (0.969+1.421+0.877)/3 = 1.089 (<SID Alvito = 1.096)

The values of SIms and SIS of Montaquila are the following:

T (LOR → SER): SI = (0+0+0.4+0+0.402+0.394) = 1.196
T^{-1} (SER → LOR): SI = (0+0+0.8+1.7+0.014+0.175) = 2.689
SIms (LOR-SER): 1.942

T (LOR → MAR): SI = (0+0+1.1+0+0.136+0.357) = 1.593
T^{-1} (MAR → LOR): SI = (0+0+0.3+0.9+0+0.359) = 1.559
SIms (LOR-MAR): 1.576

T (LOR → ID): SI = (0+0+1.2+0+0.085+0.326) = 1.611
T^{-1} (ID → LOR): SI = (0+0+0+0.7+0.126+0.941) = 1.767
SIms (LOR-ID): 1.689

T (LOR → GI): SI = (0+0+1.1+0+0.187+0.041) = 1.328
T^{-1} (GI → LOR): SI = (0+0+0.4+0+0.919+0.127) = 1.446
SIms (LOR-GI): 1.387

T (SER → GI): SI = (0+0+0.1+1+0+0.589) = 1.689
T^{-1} (GI → SER): SI = (0+0+0+0+0.812+0.527) = 1.339
SIms (SER-GI): 1.514

T (SER → ID): SI = (0+0+0+0+0.408+0) = 0.408

T^{-1} (ID → SER): SI = (0 + 0 + 0 + 0 + 0.191 + 0) = 0.191
SIms (SER-ID): 0.299

T (SER → MAR): SI = (0.3 + 0.2 + 0 + 0 + 0 + 0) = 0.5
T^{-1} (MAR → SER): SI = (0 + 0 + 0 + 0 + 0 + 0.099) = 0.099
SIms (SER -MAR): 0.299

T (GI → ID): SI = (0 + 0 + 0 + 0 + 0.975 + 0.468) = 1.443
T^{-1} (ID → GI): SI = (0 + 0 + 0 + 1.5 + 0.390 + 0) = 1.89
SIms (GI-ID): 1.666

T (GI → MAR): SI = (0 + 0 + 0 + 0 + 0.979 + 0.34) = 1.319
T^{-1} (MAR → GI): SI = (0 + 0 + 0.1 + 0.2 + 0.304 + 1.249) = 1.853
SIms (GI-MAR): 1.586

T (ID → MAR): SI = (0 + 0 + 0 + 0 + 0.303 + 0) = 0.303
T^{-1} (MAR → ID): SI = (0.6 + 0 + 0 + 0 + 0.140 + 0) = 0.74
SIms (ID-MAR): 0.521

SIS of the speakers of Montaquila: (1.942 + 1.576 + 1.689 + 1.387 + 1.514 + 0.299 + 0.299 + 1.666 + 1.586 + 0.521)/10 = 1.247 (< SID Montaquila = 1.312)

Thus, the experimental expectation E2 was also validated. Both E1 and E2 have been validated and, more importantly, they confirm the classification of the dialects stemming from the perceptive discrimination tests, as illustrated in §5.

5. Perceptive discrimination experiment

In this section we describe the perceptive experiment carried out in order to establish a terminus *a quo*: the speakers' judgments about the degree of similarity of the corpus voices. The method follows the protocol (see §5.3) of similar investigations on perceptual distance (e.g. Gooskens 1997; Preston 1999; Gooskens & Heeringa 2004). The results enable us to build a distance matrix, on the basis of which the dialects can be classified using cluster analysis (see §5.4 and Figure 18). The goal of a cluster analysis is to identify the main groups, called clusters (Jain & Dubes 1988). There are several alternatives. We used the Unweighted Pair Group Method using Arithmetic Averages (UPGMA), because we found that dendrograms generated by this method reflect distances that better correlate with the distance matrix (see Sokal & Rohlf 1962).

5.1. Speech materials

The audio corpus used to investigate dialect distances among the four selected dialects, as perceived by local listeners, is formed by the first 40 seconds of the recordings by 4 speakers from Alvito, Montaquila, Veroli, and Formia, already exploited for the formant analysis. The speakers are EL (Alvito), LOR (Montaquila), AZ (Veroli), AN (Formia).

This audio corpus was presented to local listeners during a listening experiment.

5.2. Listeners

The listeners were 4 groups of university students, one group from each location where the 4 dialects are spoken. Each group comprised 4 to 6 listeners (with a mean of 5; mean age 21.8 years, 63% females and 37% males). On average, these listeners had lived in the location in question for more than 20 years. Two of the 22 listeners (9%) said that they never speak the dialect, while the rest did so always (58%), often (21%), or seldom (12%). A large majority of the listeners (89%) had one or two parents who also spoke the dialect.

5.3. Procedure

The recordings of the four dialects were presented in randomized order on the web platform YouTube. Between every two recordings there was a 5-second pause.

After listening to the dialects, the listeners were asked to judge each dialect on a scale from 1 (similar to one's own dialect) to 10 (not similar to one's own dialect). The entire experiment lasted approximately 10 minutes and was followed by a questionnaire asking about the listeners' individual characteristics, such as language background, age, and sex.

5.4. Results: distances and classification

The mean perceptual distances among the four dialects are presented in Table 17. Each group of listeners (on the y-axis) judged the linguistic distances between the 4 dialects, including their own (on the x-axis). In this way, we obtain a matrix with 4 * 4 distances.[7]

There are two mean distances between each pair of dialects. For example, the distance that the listeners from Alvito perceived between their own dialect and the dialect of Veroli (mean judgment is 4.75 = 7.70 + 1.81/2) is different from the distance as perceived by the listeners from Veroli (mean judgment is 3.94 = 6.19 + 1.69/2). This asymmetry in perceptive measurements depends on the fact that the judgments are subjective; moreover, they are based on few observers and thus more exposed to fluctuations. For this reason, the average between the (mutual and possibly asymmetric) measurements of a speaker of dialect x and one of dialect y can be used to ‘normalize’ the outcomes.

		Dialects			
		Alvito	Montaquila	Formia	Veroli
Listeners	Alvito	1.81	1.89	4.17	7.70
	Montaquila	1.95	1.76	5.80	6.99
	Formia	3.84	4.00	1.52	7.21
	Veroli	6.91	6.01	7.58	1.69

Table 17. Mean pairwise perceptual distances of 4 dialects as perceived by 4 groups of listeners (scale range: from 1 = similar to one’s own dialect, to 10 = not similar).

The average of two mean distances is used to classify different varieties (for example, the average of the distance between Alvito-Montaquila and Montaquila-Alvito is used). In particular the perceptive distances are:

Alvito-Montaquila: (1.81 + 1.89 + 1.95 + 1.76)/4 = 1.85
Alvito-Formia: (1.81 + 4.17 + 3.84 + 1.52)/4 = 2.83
Montaquila-Formia: (1.95 + 5.80 + 3.84 + 4.00)/4 = 3.89
Alvito-Veroli: (1.81 + 7.70 + 6.19 + 1.69)/4 = 4.34
Montaquila-Veroli: (1.95 + 6.99 + 6.19 + 6.01)/4 = 5.28
Formia-Veroli: (3.84 + 7.21 + 6.19 + 7.58)/4 = 6.20

Thus the mean values are:
Alvito-Montaquila: (1.81 + 1.89 + 1.95 + 1.76)/4 = 1.85
Formia-Alvito and Formia-Montaquila: (2.83 + 3.89)/2 = 3.36
Veroli-Formia and Veroli-Montaquila and Veroli-Alvito: (6.20 + 5.28 + 4.34)/3 = 5.27

The dendrogram (Figure 18) is obtained on the basis of the mean values of Table 17.

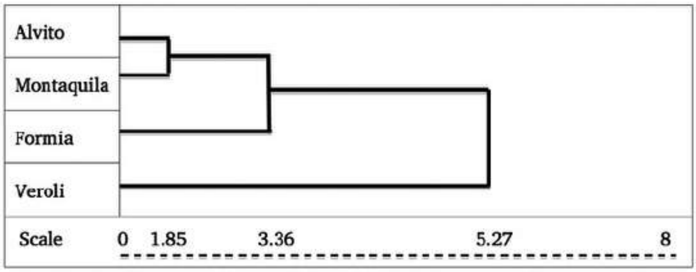

Figure 18. Dendrogram derived from the 4 * 4 matrix of perceptual distances showing the clustering of dialects. On the horizontal scale, distances are given in the scale as used by the listeners.

On the basis of the distance matrix, the dialects can be arranged in clusters comprising groups of similar dialects. Clusters may consist of sub-clusters, which in turn may consist of sub-subclusters, and so on: the result is a hierarchically structured tree. According to this classification, the dialects from Alvito and Montaquila are judged to be more similar, the dialect from Formia less similar, and the dialect from Veroli the most different.

This arrangement matches the dialectological grouping of these dialects in southern Lazio (Vignuzzi 1988; Avolio 1995; Loporcaro 2009). Most importantly for our purpose, it also matches the results of the topological analysis in §4.4, where the dialect of Veroli appears to be topologically more distant from the cluster Alvito-Formia-Montaquila.

Let us summarize the main results and draw some conclusions.

6. Main results

The model and the present dialectometric investigation have been tested on a speech corpus of four Italian dialects, obtaining a classification that matches the dialectological literature (§5.4).

Starting from the formant charts of the stressed vowels of selected speakers, a topological comparison was implemented, enabling the computation of a new kind of metrics and of a Similarity Index (SI) (§3.5). The SI relative to two speakers or two dialects was respectively defined as SIms (between a pair of speakers) and SImd (between a pair of dialects); SID (of a given dialect) was defined as the arithmetic mean among the values of all SImd concerning all dialect comparisons relative to a given dialect; conversely, SIS was defined as the arithmetic mean among the values of all SIms relative to the speakers of a given dialect (§4.2).

The mathematical relationships among these indexes and their value in order to predict the degree of similarity among dialects has been checked against two experimental expectations: E1 and E2 (§4.2). The topological analysis of the audio corpus validated both expectations, and this result matched the classification of the dialects stemming from a perceptive discrimination test (§4.4).

In order to evaluate why a topological model does represent the identities/differences among vowel systems better than the traditional approach, we refer readers to the data in the paper. For instance, according to the speakers' impressionistic judgments (§5.4), the dialects from Alvito and Montaquila are judged as the most similar in the experimental set. Note that a traditional analysis of the vowel systems would suggest – for instance – that /u/, /a/, /e/, /i/ of Montaquila are different from /u/, /a/, /e/, /i/ of Alvito, because their F_2 show very different values (Tables 5 and 11), whereas the values of other vowels are similar (for both F_1 and F_2). By contrast, the topological approach shows that – considering the stressed vowels – the dialects of Alvito and Montaquila appear very similar, and both differ from the dialect of Veroli (§4.4). Thus, the topological approach matches and predicts the speakers' judgments, whereas the traditional approach cannot provide a robust prediction (or gives disjointed predictions concerning the individual vowels).

7. Conclusions

The topological representation of the vocal space presented in this paper makes it possible to establish robust equivalence classes, which generate predictions about the degree of similarity of different formant spaces/charts (as computed on the basis of the mean values of F_1 and F_2). The model was tested on the formant charts of the stressed vowels of selected speakers of a few central Italian dialects. The predictions were successfully validated by a perceptive discrimination test.

Needless to say, this is a pilot study, and its major goal was methodological, rather than heuristic. Further verifications should be implemented on larger corpora, in order to further validate the model.

Such topological representations could also prove useful in forensic studies, and may also provide a framework for the study of voice mimickers, comparing and discriminating different imitations with respect to the imitated voice.

Finally, this study can provide a topological interpretation of Flege's Speech Learning Model (SLM), at least relative to stressed vowels. Flege's model (Flege 1987, 1988, 1995, 2002; Flege, Munro & Fox 1994; Flege, Schirru & MacKay 2003) assumes that the phonetic categories of L1 and L2 of a bilingual speaker exist in one and the same phonological space, in which mutual influence between the speech sounds of both languages is predicted. More specifically, such model hypothesizes that similar sounds of L1 and L2 are linked to one another in a single representation, on the basis of the individual's perception. For such cross-language associated sounds, called diaphones, the model predicts category assimilation; that is, the sounds come to resemble one another over time. The present topological framework may be able to trace such transformations, for instance those that lead the stressed vowels of two languages to resemble one another over time.

Abbreviations

A: polygon A.
B: polygon B.
Bark: a frequency scale on which equal distances correspond with perceptually equal distances.
DTP: D'Arcy Thompson's Pictures (cf. §3.1).
E1: experimental expectations concerning the value of SImd.
E2: experimental expectations concerning SIS < SID.
ERB: Equivalent Rectangular Bandwidth. It is a psychoacoustic measure, which gives an approximation to the bandwidths of the filters in human hearing, using the simplification of modeling the filters as rectangular band-pass filters.
FFT: Fast Fourier Transform. It converts a signal to a representation in the frequency domain, or vice versa
LPC: Linear predictive coding. It is a method of encoding signals according to which the value of a signal at a specific point in time is a linear function of the past values of the quantized signal.
Mel: a perceptual scale of frequency values judged by listeners to be equal in distance from one another.
MFCC: Mel-Frequency Cepstrum Coefficients.
σ: Standard Deviation.
SI: Similarity Index.
SID: arithmetic mean among the values of all SImd concerning all pairs of dialects in which a given dialect occurs.
SIS: arithmetic mean among the values of all SIms concerning all speakers of a given dialect.
SImd: arithmetic mean between the values of SI of a pair of dialects.
SIms: arithmetic mean between the values of SI of a pair of speakers.
T: transformation.

Notes

[1] Several studies have tried to eliminate speaker-related variation by designing procedures categorized under the heading of 'vowel normalization'. Formant measurements show considerable variation related to anatomical/physiological differences between speakers. The discrepancy between how listeners are affected by differences between speakers and how these differences affect formant measurements is also known as the 'lack of one-to-one correspondence between acoustics and perception'. Many studies have sought to improve the correspondence between the acoustic and perceptual dimension of speech. One of the hypotheses is that listeners naturally perform normalization when perceiving speech sounds. Humans are assumed to somehow perceptually 'even out' differences between speakers. Therefore, research in the acoustic domain aims to understand the normalizing behavior of listeners from two perspectives. The first one (formant-based approaches) represents the formant frequencies on auditory scales (e.g. Bark-scale, Mel-scale, ERB-scale). The goal of the second perspective (whole-spectrum approaches) is to understand the perceptual and cognitive processes involved in vowel processing. To this end, various normalization procedures have been developed to serve as stages in psychological models for human vowel recognition. The primary purpose of these procedures is to model human speech perception in order to explain how listeners categorize vowel sounds. Generally speaking, these approaches use spectral information beyond the center frequency of the spectral peaks used in formant-based approaches. Whole-spectrum approaches assume that all spectral information is relevant, and that no speaker-specific information should be discarded. For instance, these approaches use MFCC (Mel-Frequency Cepstrum Coefficients).

Limited to the acoustic data plotted in a formant chart, our approach tries and achieves the same normalization, since its goal is to sort predictions about the listeners' perceptive judgment. Similar to normalization procedures, it aims at filling the gap between acoustics and perception. Therefore, normalization procedures and topology converge in their goal, and this is why we did not apply normalization to the formants data. However, in order to provide evidence that our topological transformations are not affected by the scale used to measure the formants, in Appendix 3 we compare the same data as plotted in Hz, Bark, and Mel. As one can see, the formant chart polygons remain unchanged and so are the transformation equations.

Further research may weigh the power of resolution of the topological approach against MFCC or other normalization practices. Note, at any rate, that our corpus consists of a small number of female voices, with values of variance and standard deviation added to the mean formant measurements. Thus, considering the homogeneity and statistical significance of our data, they are hardly in need of normalization.

[2] The model does not stem from linguistic theory but from mathematics, geometry, and topology. The main idea is that a formant chart is both a linguistic and a mathematical object; thus, its mathematical properties may involve and engage some linguistic effects. Specifically, this is proved by the fact that the topological predictions of the model match the perceptive judgments of the speakers. In the end, from an operational point of view, we believe that a topological model can become a linguistic model on condition that its predictions agree with the actual linguistic behavior. From an epistemological point of view, the possible motivation of the relationship between topology and linguistics may be understood as a particular emergence of the more general topological properties of the human cognitive system, as accounted for by neurosciences (see for instance Bullmore & Sporns 2009, Bullmore & Bassett 2011, Sizemore *et al.* 2018).

[3] The web page describes the HyperCard stack called DTP. However, the 'SheepShaver' classic Mac emulator must be installed in order to enable DTP to run under emulation on a recent Mac.
[4] The numerical implementation of x and y in the default-T corresponds to the default settings of the user interface of DTP, that is to a DTP with no T.
[5] Further details on data collection and specifics on the analysis of the linguistic materials themselves etc., can be drawn from the website of the 'Centro': <linguistica.unicas.it/dlm/doku.php>.
[6] Moreover, in the case of Italian dialects, the unstressed vowels often are allophones of the stressed vowels. Therefore, unstressed vowels may be regarded as contextual realizations (allophones) of the stressed ones, i.e. they are in complementary distribution with them: it could be impossible to plot unstressed and stressed vowels in the same formant chart. In other languages, where unstressed vowels are true phonemes, it would be necessary to also analyze the formant chart of unstressed vowels.
[7] Our goal was to investigate the dialect distances between dialects as perceived by the listeners, in particular the distance between their own dialect and the other dialects under investigation. Thus, we needed to measure two points: how each listener rates his own dialect and how s/he rates the other dialects under investigation. That is why the listeners are asked to compare the dialects to their own variety in order to generate general dialectal distances.

Bibliographical References

Abercrombie, David 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Avolio, Francesco 1995. *Bommèsprə. Profilo linguistico dell'Italia centro-meridionale*. San Severo: Gerni.

Bradlow, Ann R. 1993. A cross-language comparison of vowel production and perception: language-specific and universal aspects. *Working Papers of the Cornell Phonetics Laboratory* 8. 29-85.

Bullmore, Edward T. & Sporns, Olaf 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10. 186-198.

Bullmore, Edward T. & Bassett, Danielle S. 2011. Brain Graphs: Graphical Models of the Human Brain Connectome. *Annual Review of Clinical Psychology* 7. 113-140.

Fant, Gunnar 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fant, Gunnar 1966. A note on vocal tract size factors and non-uniform f-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1. 22-30.

Fant, Gunnar 1973. Acoustic Description and Classification of Phonetic Units. In Fant, Gunnar, *Speech Sounds and Features*. Cambridge, Mass.: The MIT Press. 32-83.

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47-65.

Flege, James E. 1988. The production and perception of foreign language speech sounds. In Winitz, Harris (ed.), *Human communication and its disorders: a review*. Norwood, NJ: Ablex. 224-401.

Flege, James E. 1995. Second language speech learning: theory, findings, and problems. In Strange, Winifred (ed.), *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, James E. 2002. Interactions between native and second-language phonetic systems. In Burmeister, Petra; Piske, Thorsten & Rohde, Andreas (eds.), *An integrated view of language development: papers in honor of Henning Wode*. Trier: Wissenschaftlicher Verlag. 217-244.

Flege, James E.; Munro, Murray J. & Fox, Robert A. 1994. Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America* 95,6. 3623-3641. <dx.doi.org/10.1121/1.409931>.

Flege, James E.; Schirru, Carlo & MacKay, Ian R.A. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40,4. 467-491.

Goldstein, Louis 1984. Vowel Shift and Articulatory-Acoustic Relations. In Van den Broecke, Marcel Peter René & Cohen, Antonie (eds.), *Abstracts of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris. 267-273.

Gooskens, Charlotte 1997. *On the role of prosodic and verbal information in the perception of Dutch and English language varieties*. Doctoral dissertation. University of Nijmegen, the Netherlands.

Gooskens, Charlotte & Heeringa, Wilbert 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change* 16. 189-207.

Jain, Anil K. & Dubes, Richard C. 1988. *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice Hall.

Ladefoged, Peter 1967. *Three areas of experimental phonetics*. London: Oxford University Press.

Laver, John & Trudgill, Peter 1991. Phonetic and linguistic markers in speech. In Laver, John (ed.), *The Gift of Speech: Readings in the analysis of speech and voice*. Edinburgh: Edinburgh University Press. 235-264.

Lindau, Mona 1978. Vowel Features. *Language* 54. 541-563.

Liu, Chang & Kewley-Port, Diana 2004. Vowel formant discrimination for high-fidelity speech. *Journal of the Acoustical Society of America* 116,2. 1224-1233.

Loporcaro, Michele 2009. *Profilo linguistico dei dialetti italiani*. Bari: Laterza.

Oglesbee, Eric & Kewley-Port, Diane 2009. Estimating vowel formant discrimination thresholds using a single-interval classification task. *Journal of the Acoustical Society of America* 125,4. 2323-2335.

Potter, Ralph K. & Steinberg, John C. 1950. Toward the specification of speech. *Journal of the Acoustical Society of America* 22. 807-820.

Preston, Dennis 1999. *Handbook of perceptual dialectology*. Amsterdam: Benjamins.

Rosner, Brian S. & Pickering, John B. 1994. *Vowel Perception and Production*. New York: Oxford University Press.

Sizemore, Ann E.; Giusti, Chad; Kahn, Ari; Vettel, Jean M.; Betzel, Richard F. & Bassett, Danielle S. 2018. Cliques and cavities in the human connectome. *Journal of Computational Neuroscience* 44,1. 115-145. <doi: 10.1007/

s10827-017-0672-6>.
Sokal, Robert R. & Rohlf, F. James 1962. The comparison of dendrograms by objective methods. *Taxon* 11. 33-40.
Thompson D'Arcy, Wentworth 1917. *On Growth and Form*. Cambridge: Cambridge University Press.
Vignuzzi, Ugo 1988. *Italienisch: Areallinguistik VII. Marche, Umbrien, Lazio. Aree linguistiche VII. Marche, Umbria, Lazio*. In Holtus, Günter; Metzeltin, Michael & Schmitt, Christian (eds.), *Lexicon der romanistischen Linguistik*. Vol. 4. Tübingen: Niemeyer. 606-642.