DOI: 10.17469/O2109AISV000003

CLAUDIA CROCCO, BARBARA GILI FIVELA, GIUSEPPE MAGISTRO

Comparing dialectal and Italian prosody: the case of Venetian

The following paper aims at setting out a novel methodology in the prosodic comparison between two varieties in contact, the dialect spoken in Venice and the regional Italian spoken in Venice. By deploying a reading task, we compare the rhythmical properties of the two systems and review different metrics. We show that speakers can switch their metrical organization when switching language, but this is sensitive to those segmental processes which differentiate the two systems in contact.

Keywords: contact, rhythm, vanishing "l", "elle evanescente".

1. *Introduction*

The present study aims at comparing dialect and Italian from the prosodic point of view, with a primary methodological goal: we aim at testing a set of suitable procedures to identify and possibly quantify prosodic differences between dialectal and Italian varieties.

As a case-study, we consider Venetian Italian (VI) and urban Venetian dialect (VD). VD is an Italo-Romance vernacular or primary dialect (Coseriu, 1981), i.e., a sister language of Tuscan, from which Italian stems from (Serianni, Trifone, 1993). The comparison presented in this article, therefore, involves two historically related and yet grammatically and phonologically distinct linguistic systems that are in long-standing contact and co-exist in the city of Venice as varieties widely spoken in everyday conversation (Berruto, 2012, Ferguson, 2007). Compared to other Italo-Romance dialects, VD has enjoyed a certain prestige both in the past and today (Cortelazzo, Paccagnella, 1992, Dal Negro, Vietti, 2011). Accordingly, dialect-standard bilingualism is widespread in the region and in the city of Venice, and dialectal speech does not suffer a social stigma (Dal Negro, Vietti, 2011, ISTAT 2017).

VD is characterized in its pronunciation by the so-called *cadenza* (or *calada* in VD; Ferguson, 2007), i.e., a sing-song rhythmical cadence, a feature hinting to prosodic properties of VD partially diverging from those ascribed to Italian varieties (Gili Fivela, Avesani, Barone, Bocci, Crocco, D'Imperio, Giordano, Marotta, Savino & Sorianello, 2015). Magistro and Crocco (2022) explored the rising movements characterizing the final stretch of statements in Veneto dialects, proposing that they may play a role in

the *calada*. In this paper, we focus on durational differences between VD and VI and examine their possible impact on the rhythmic organization of the languages at stake.

The paper is structured as follows: in § 2, we discuss the link between durational variation and rhythm organization. In § 3, several relevant phonetic and phonotactic features of VD are presented. § 4 is dedicated to the methodology adopted to collect (§ 4.1) and pre-process the dialectal and Italian datasets (§ 4.2), with particular attention to the procedure adopted in specific cases (§ 4.2.1). Subsequently, we present the results of a statistical analysis of the durational measurements (§ 5) and formulate hypotheses about the possible source of the observed durational variation (§ 5.1). After exploring these hypotheses (§ 6), we discuss the implications of the results and draw the conclusions of the study (§ 7).

2. *Levels of rhythm organization*

Along the lines of Clarke (1999), Kohler (2009) and Arvaniti (2009), a distinction can be made between timing and rhythm: the former has to do with the duration of events (i.e., durational variability), while the latter regards the regular pattern extracted by the listener from, a.o., durational features. Although linguistic rhythm cannot be reduced to durational variability (Arvaniti, 2009), durational patterns are likely to be relevant for the organization and the perception of rhythm, as this phenomenon unfolds over time (Turk, Shattuck-Hufnagel, 2013).

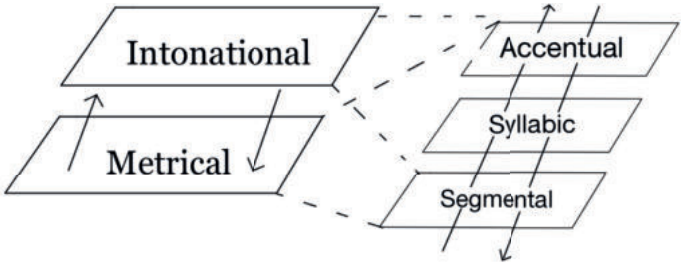
Languages may differ in their prominence-marking strategies (Andreeva, Barry & Koreman, 2014) and, therefore, in their use of duration to mark stressed syllables. In languages such as Italian, for instance, duration is essentially linked to prominence, as it is the most stable correlate of lexical stress both in production and perception (Bertinetto, 1981, Krämer, 2009, D'Imperio, Rosenthall, 1999). On the other hand, lengthening of stressed syllables is far less evident in Spanish than in Italian (Alfano, Savy & Llisterri, 2009; White, Payne & Mattys, 2009; see also Schmid, 2014), and differences are reported in relation to varieties of Italian as well, as prominence-related lengthening seems far more evident in Sicilian than in Venetian Italian (White et al., 2009).

Further, rhythmic differences across languages concern both low-level differences in temporal organization, and high-level differences involving, e.g., the spacing of stresses and accents or the tolerance to arrhythmic configurations (Jun, 2012, 2014, Falk, Rathcke, Dalla Bella, 2014, Frota, Moraes, 2016, Arvaniti, 2007, 2009, Turk, Shattuck-Hufnagel, 2013). Rhythmic analysis should therefore include the interaction between such layers and try to address the question how exactly the low-level temporal features are related to the high-level rhythmic organization, i.e., to the prominence hierarchy and its implementation (Fletcher, 2010, Rathcke, Smith, 2015), and the other way around, that is how the high-level organization, e.g., spacing of accents, is related to the rhythmic differences. In line with this wide perspective on the issue, in this paper we also try to trace back the possible sources of the observed temporal variation at the segmental level, by linking the experimental

results to what is known about the phonetics and phonology of the languages at stake (cf. Turk, Shattuck-Hufnagel, 2013).

Models such as the coupled oscillators proposed by Barbosa (2002, 2007) try to account for the interaction between prominence patterns governed by higher linguistic levels (phrase stress oscillator) and the syllabic sequence organized around vowel onsets (syllabic oscillator). In this model, cross-linguistic differences concerning vowel and consonant reduction are specified in the gestural lexicon. This intrinsic, language-specific level of timing interacts with the prosodic organization to produce actual segmental durations. However, segmental adjustments involving a.o. vowel lengthening, vowel insertion etc. can also be driven by the need to expand the text to provide more site for the tune realization (Grice, Savino, Roettger 2018, Roettger, Grice 2019). Therefore, just as the accentual pattern can be adapted to the words, syllables and segments composing the text, as it happens in well-known cases of tonal truncation or repulsion, adjustments such the insertion or the lengthening of a vowel can also be induced in the text by the tune (see Fig. 1). This suggests that timing effects, while being indeed language-specific, may be not as such independent from the high-level rhythmic organization.

Figure 1 - *Interactions between metrical and tonal tiers and between high-level and low-level components of rhythm*



2.1 Measurements of consonant-vowel ratio

Starting from Dauer's (1983, 1987) observation that vowel reduction may play a role in the perception of rhythm, several scholars have tried to quantify the consonant-vowel ratio by means of specifically developed metrics, in order to assign languages to rhythm classes (a.o. Dellwo, 2004, Dellwo, Wagner, 2003, Grabe, Low, 2002, Ramus, Nespor & Mehler, 1999). Such metrics provide different measures under the common assumption that the consonant-vowel ratio is a direct reflection of the language rhythm organization. The first rhythm metric was proposed by Ramus, Nespor & Mehler in 1999. This metric is based on some of the claims made by Dauer (1984) concerning possible phonetic and phonological correlates of stress-timed languages. The Deltas calculate the standard deviation of vocalic intervals (ΔV), the standard deviation of consonantal or intervocalic intervals (ΔC) and the percentage of vocalic intervals (%V). In the framework of the rhythm classes hypothesis, Ramus and colleagues hypothesize that the duration of vocalic and consonantal intervals

would show a stronger variation in stress-timed than in syllable-timed languages. Accordingly, higher values of ΔV and ΔC are expected in stress-timed languages compared to syllable-timed languages. Additionally, syllable-timed languages, which are characterized by less complex consonant clusters, would present a higher vocalic percentage %V compared to stress-timed languages. Since Deltas are extremely sensitive to speech rate variations, Dellwo and Wagner (2003) and Dellwo (2006) tried to improve the metric by normalizing data for speech rate. The normalized Deltas are called Varcos. Parallel to the Deltas, higher values of VarcoC and VarcoV are expected in stress-timed compared to syllable-timed languages.

A further metric is the so-called Pairwise Variability Index (PVI), proposed by Grabe and Low (2002), originally conceived to grasp the timing differences between closely related dialects of English. The PVI differs from Deltas and Varcos as it also considers the temporal sequence of vocalic and consonantal intervals. The formula of the row PVI (rPVI) computes the difference in duration between one interval and the following in a pairwise fashion, and then calculates the average of all differences. Since vowels are expected to be more sensitive to speech rate variations, Grabe and Low (2002) propose a nPVI (normalized PVI) for the calculation of vocalic intervals. As in the cases of Deltas and Varcos, also for the PVI stress-timed languages are expected to show lower values of rPVI and nPVI than syllable-timed languages.

Finally, a different metric has been proposed by Bertinetto & Bertini (2008). This metric differs substantially from other previously proposed in that it introduces a phonological dimension in the quantification of durational facts and distances itself from the stress-timing/syllable-timing dichotomy. The Control and Compensation Index (CCI) is a modification of the rPVI proposed by Grabe and Low (2008). In the CCI, the duration of each vocalic or consonantal interval is divided by the number of phonological segments included in the interval. Accordingly, geminate consonants and phonologically long vowels count as two segments. The CCI represents the level of "compression" allowed in a language/variety, i.e., the extent to which vocalic and consonantal segments can be lengthened or shortened in the context where they occur. Differences in the level of compression account for differences across languages; according to the authors' hypothesis, controlling languages allow for a low level of compression, whereas compensating languages allow for a high level of compression. Considering a space organized along the two dimensions of vocalic control and compensation (VCCI), and consonantal control and compensation (CCCI), controlling languages are expected to be scattered along the bisector, whereas compensating languages are expected to cluster below the bisector, in the lower right quadrant of the chart.

However, rhythm metrics as an instrument to assign languages to rhythm classes has received strong criticism, along with the rhythm class hypothesis itself (Kohler, 2009, Arvaniti, 2009, 2012, Rathcke, Smith, 2015). While the metrics provide relatively consistent results when applied to prototypical cases of stress-timed and syllable-timed languages, such as English and Spanish, they fail to reliably assign other languages, such as Greek or Thai to a specific rhythm class. Furthermore, the

results obtained through the metrics are heavily influenced by the characteristics of the analyzed data, in terms of speech rate, speech style, segmental composition etc., and consequently have a weak predictive power (Arvaniti, 2009, 2012). These results, besides indicating that rhythm metrics are not reliable as a tool to assign a language to a given rhythm class, also cast doubts on the validity of the rhythm classes hypothesis itself. Additionally, since the same durational measurements can derive from different sources in different languages, an additional shortcoming of such measurements is represented by their inability to identify the source of the observed variation (Arvaniti, 2009, Turk, Shattuck-Hufnagel, 2013).

In our view, the arguments put forward by the above-mentioned scholars are compelling. Rhythm metrics only provide a rough measurement of the durational facts, that are not *per se* typologically or phonologically relevant. Moreover, durational variation can derive from different sources, and is not the sole phonetic exponent of linguistic rhythm (Turk, Shattuck-Hufnagel, 2013). Such measurements, in fact, simply characterize the durational organization of a speech sample in terms of vocalic and consonantal intervals. Keeping this in mind, in this study we approach this set of measurements in a way that decidedly diverges from the preceding studies, using them to explore the hypothesis that one and the same bilingual speaker of VD and VI can adapt his/her durational organization when speaking Italian vs. dialect. Therefore, we use the rhythm metrics described above (Deltas, Varcos, PVI and CCI; implemented in the program *Correlatore*, Mairano, Romano, 2010) exclusively as a tool to quantify temporal differences between VD and VI at the segmental level in terms of vocalic and consonantal intervals; in principle, other measurements could have been used too. Accordingly, we will not frame the results in the rhythm classes hypothesis.

3. *Venetian Dialect: Phonetics and phonotactics*

Although systematic studies on VD phonology are still missing, the available research indicates that VD and Italian differ in several aspects that can be relevant for their rhythmic and prosodic organization. The phonotactics of VD, in particular, is fairly simple when compared to that of other northern dialects such as Romagnolo. Venetian presents 24 syllabic types (Schmid, 2014), most of which are in common with Italian (Schmid, 1998). Previous measurements on VD based on the rhythm metrics (Schmid, 2014), indicate for this dialect a relatively high proportion of vowels (%V) compared to other Italo-Romance dialects, and relatively low variability of the consonantal and vocalic durations (low standard deviation values for consonantal and vocalic duration, ΔV and ΔC). Overall, Schmid's (2014) measurements place VD in the area of syllable-based dialects.

VD has a number of morphological-phonological properties that are absent in Italian, e.g., final vowel/syllable apocope in several contexts and the so-called *'l' evanescente* (vanishing 'l'), i.e., elision/approximant realization of intervocalic /l/ (Zamboni, 1988). Such properties affect the way syllables are realized in VD

with respect to Italian. As for apocope, while Italian allows reduction and re-syllabification within and across words in informal, hypo-speech contexts, in VD re-syllabification phenomena show a (more) systematic character and can lead to fixed, univerbated forms (cf. Ferguson, 2007):

(1) *nol* < *no* + *el* "not the",
pel < *per el* "for the",
naltra < *un'altra* "another" [fem.],
chel < *che el* "that the"

Such cases provide relevant indications of possible differences in the syllable count between Italian and VD in comparable utterances.

A specific issue in VD is represented by the so-called *'l' evanescente* (see e.g., Tomasin, 2010), which is one of the three allophonic variants of the phoneme /l/ in this dialect. According to Tomasin (2010), such variants are:

- [l] in pre- or post-consonantal position, as in *folpo* 'octopus', *cantarla* 'to sing it';
- [ɬ] ([ɻ] in Tomasin's transcription) initial and intervocalic position, excepted when one of the vowels is a palatal. Such *'l' evanescente* is described as an approximant («approssimante dorsopalatale rilassata»), as in [gondoła] 'gondola';
- finally, /l/ is canceled ("dileguo") in intervocalic position when it precedes or follows a palatal vowel, as in *fià̀r* 'spin' or *vea* 'sail'.

In the Italo-Romance domain, l-vocalization in intervocalic position is attested in several northern and southern dialects, in which variants including [ɬ], [j] and deletion are possible in different contexts (Rohlfs, 1966: 305ff.). Rohlfs' data suggest that this development is driven by the presence of a palatal vowel, which will then diverge from the development of /l/ into the labio-velar [w]. In Veneto, the approximant realization of /l/ and its deletion are likely to be relatively recent developments, since they are not attested in ancient text and they also lack in Goldoni's language (Rohlfs, 1966: 308), which may be considered as a most prominent example of literary use of the variety. The feature seems to spread from Venice to other varieties spoken in the region (e.g., Paduan; Tomasin, 2010: 731).

To the best of our knowledge, phonetic correlates of l-vocalization in intervocalic position have not yet been experimentally investigated in Romance. Experimental research on the realization of lateral consonants mostly concerns the case of l-velarization in coda position (Recasens, 2012). As for Veneto, the current account of l-vocalization is mostly based on descriptive studies such as Lepschy (1962) and Zamboni (1988), while experimental investigations of the phenomenon are still missing.

The allophonic realization of /l/ as [ɬ] or its deletion represent a potentially relevant issue when it comes to the temporal and, more generally, to the prosodic comparison between regional VI and VD. Although it can be argued that non-syllabic [ɬ] functionally acts as a consonantal incipit, its phonetic realization seems indeed *vocalized*, possibly contributing durational differences in vocalic and consonantal intervals in the two relevel varieties.

4. *Methodology*

In order to verify if bilingual speakers can switch temporal organization when speaking Italian vs. dialect, we measured segmental durations in two sets of VD and VI utterances closely comparable from the lexical, syntactic and informational point of view, and compared durational variability *within each individual* examined for the study. We first identified the metrics that better distinguish between VI and VD using statistical methods. Subsequently, we interpreted the results of these metrics against the background provided by the available knowledge of the phonetics and phonology of VD and Italian. In particular, we focused on the impact of segmental differences on syllable realization and syllable count.

4.1 Data collection

We recorded 5 bilingual speakers reading a set of dialogues corresponding to question-answer pairs. The recordings were taken at 44100Hz 16-bit *wav* format with a Blue Yeti microphone with the monodirectional polar pattern. All speakers declared a high level of proficiency in both Italian and Venetian, as well as claiming to use dialect on a regular basis. Speakers were aged 50-75 and coming from the neighborhood of Castello, in the historical center of Venice. The experiment consisted in two recording sessions (separated at least one day from the other). In the first session, speakers were prompted to read dialogues in Venetian or Italian, and in the second one, they were asked to record the other language. The order of the blocks was presented randomly to avoid potential saturation biases. Although the analysis of natural speech represents the ideal goal to aim for, we chose to rely on read speech for this study to control the experimental layout. This choice is due to the fact that durational measurements hare sensitive to segmental composition and speech style. Given the exploratory nature of the study, it was necessary to start building hypotheses with "clean data", without having to disentangle complicating factors. Furthermore, a rigid segmental layout is an offset to the limited number of recorded speakers. The experimental corpus consisted in a set of 15 target dialogues, eliciting three types of declaratives, 5 Broad Focus declaratives (1), 5 Contrastive Focus (2), and 5 Narrow-Informational focus (3). These dialogues had the same meaning across languages and were designed to keep as much similar segmental layout as possible. For example, the syllable count was kept constant across languages, together with lexical items. As appreciable in the reference examples, each target sentence was made of a bi-syllabic verb followed by the article '*la*' and a trisyllabic noun starting with nasal and having lexical stress on the antepenultimate syllable (Fig. 2). Besides 15 target dialogues, an equal number of fillers was elicited.

Table 1 - *Example of sentences*

Italian	Venetian	Translation
Broad Focus		
A: *Cosa fai stasera?* B: **Cucio la manica.**	A: *Cossa ti fa de sera?* **B: Cuzo la manega.**	'What will you do tonight?' 'I will sew the sleeve'
Contrastive Focus		
A: *Cuci il bottone stasera*? **B: Cucio la manica, stasera.**	A: *Ti te cuzi 'l boton stasera*? **B: Cuzo la manega, stasera.**	'Will you sew the button today?' 'I will sew the sleeve, tonight'
Narrow Informational Focus		
A: Cosa cuci stasera? **B: Cucio la manica, stasera.**	*A: Cossa ti cuzi stasera?* **B: Cuzo la manega, stasera.**	'What will you sew tonight?' 'I will sew the sleeve, tonight'

4.2 Data pre-processing

The target sentences were manually cut in Praat (Boersma, Weenink, 2022). Thereafter they were automatically segmented using the Forced Aligner MAUS (Schiel, 1999) trained for Italian. The intervals were sanity-checked by means of the following procedure: a small sample of 10 TextGrids was visually and auditorily inspected in Praat by each author. By joint comparison of the corrected TextGrids, the major issues of the automatic alignment within the corpus were discussed and debugged. Once the guidelines of manual correction were outlined, each author corrected another subsample of 40 sentences. Mutual annotation agreement was calculated using Intraclass Correlation Agreement (ICC) using the package *irr* in R (Gamer, Lemon, Fellows & Singh, 2019). After checking the main statistical assumptions, the ICC test was performed. The choice of the test was driven by the continuous nature of time-aligned segmentation: while K coefficient is well-suited for categorical variables, ICC provides a valid statistic for continuous dimensions. Specifically, a two-ways random ICC test for absolute agreement was performed and an F value $F(920, 1840) = 10{,}7$ with ICC = .76, $p < 0{,}001$, indicating substantial agreement on boundary placement. Once ascertained the inter-annotator agreement, each author corrected individually an equal part of remaining items (33 items per author = 99 + 1+ 50 jointly corrected = 150). The rationale behind this complex procedure is threefold. First, it alleviates the task of manual segmentation of the phonemes, where only sanity-check must be performed by researchers. Second, by examining the 30% of the automated output in distinct steps, it is possible to ensure that the human intervention on the data was consistent. Finally, a reliable segmentation is essential when computing metrics from duration values as we intended to do.

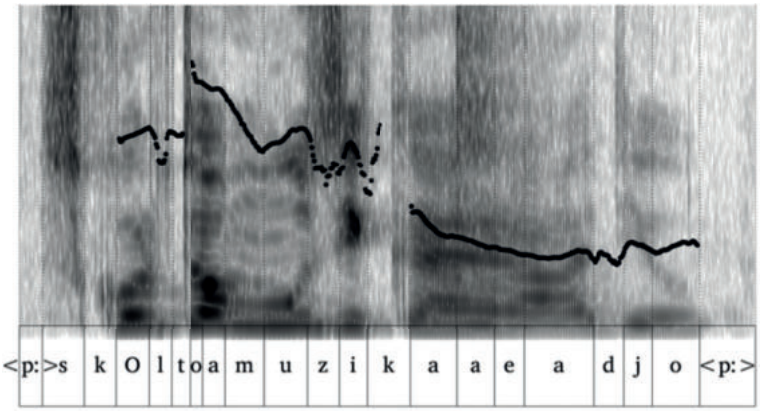
4.2.1 Annotation procedure for *'l' evanescente*

As discussed in § 2, /l/ in VD can be realized as [l], [ɬ], or can be deleted, according to the phonotactic context. In line with the literature (see § 3), we expect /l/ to be realized as [ɬ] in the target sentences, i.e., as an approximant consonant in articles (*la*, "the") and prepositions (*ala*, "to the").

For the annotation of the *'l' evanescente* we adopted the following procedure during the manual check of the Forced Aligner MAUS segmentation (Schiel, 1999): the annotator assessed the realization of the article *ea* "the" by listening to the prosodic phrase ending with the target word. In several cases, listening and instrumental inspection indicate cancellation of /l/ also in contexts in which [ɬ] is expected. In such cases, no allophone of /l/ was reported in the segmentation. An illustration of multiple l-cancelation is the utterance presented in Fig. 2.

Since the metrics implemented in the *Correlatore* compute the V-C proportions, the attribution of a vocalic or consonantal status to /l/ is likely to affect the result of the comparison between Italian and dialect. To avoid introducing biases in the comparison, we adopted a conservative approach by creating two copies of the Venetian dataset to feed the *Correlatore*, the first with /l/ labeled as lateral consonant [l], and the second with /l/ labeled as vocalic [e]. When the syllabic incipit /l/ was absent, only the syllabic nucleus was segmented and annotated, and the same TextGrid (with no interval corresponding to /l/) was used to feed the *Correlatore*.

Figure 2 - */l/ vocalization in the utterance Scolto la musica ala radio "I listen to music on the radio" (female speaker; 3vecfmus)*



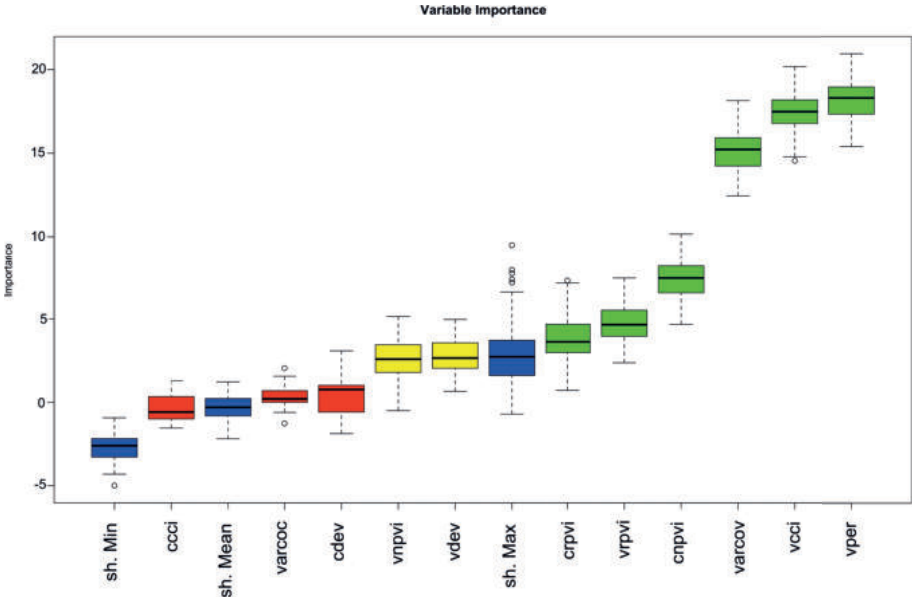
5. *Results: Rhythmic measurements*

5.1 Which dimension matters?

The 150 corrected TextGrids were batch-processed in *Correlatore* (Mairano, Romano, 2010) to extract the relevant rhythmic measures. The metrics were then exported in R, where they were tested to evaluate their capability in distinguishing between VI and VD. To this end, a feature selection algorithm was applied. We implemented a Boruta algorithm using the homonymous package in R. Boruta is a Random Forest-based algorithm, which compares the importance of each feature of the dataset with irrelevant randomly mixed features (the so-called *shadows*) to evaluate which real features are relevant for the classification (Kursa,

Rudnicki, 2010). Boruta, in comparison with other feature selection algorithms, is advantageous for our case: it is a method which does not seek to restrain the dimensionality to the minimum, i.e., it does not take out unnecessary features when others already do the job. This is a desideratum, because the algorithm will not penalize redundant features, which are naturally present in the metrics (the metrics all depend on the length of vowels and consonants, but with different formulas). We report in Fig. 3 the output by Boruta.

Figure 3 - *Results of feature selection with a Boruta algorithm*



The blue boxplots represent the Shadow features, used by Boruta as baseline. The red features are irrelevant for the classification VD-VI, while the yellow ones have slightly more chance to be relevant. The green ones, on the other hand, represent the most relevant features. In our paper, we will focus on the three most important ones, that is %V, VarcoV and VCCI. For the reader's convenience, we recall here that these measurements refer, respectively, to the percent of vocalic intervals in the measured speech (%V), the normalized standard deviation of vocalic interval duration divided by the mean (VarcoV), and the average duration of each vocalic interval divided by the number of phonological segments included in the interval (VCCI).[1] Of course, the algorithm cannot tell us anything about the direction

[1] We also checked whether there is an interaction between the sentence type (broad, contrastive and narrow focus) and the variety in the rhythmic distribution. This was performed with a clustering algorithm, kNN (see next section). The results showed a scattered distribution where the dimension of sentence type cannot represent a grouping factor accounting for the observed variability. We then continue to explore the effects of the continuous rhythmic indexes spotted by Boruta. We will test again whether the pragmatic condition has ultimately an effect in § 5.3.1.

of the relationship between the categorical dependent variables and the relevant features. In other words, while the algorithm helps us to identify the most useful features to distinguish VI from VD, it does not provide information about which of the two varieties has e.g., a higher %V. Next sections will aim at covering this gap.

5.2.1 Clustering methodology: kNN

In order to capture the directionality of the relevant features, we used a k-Nearest Neighbour Machine Learning algorithm (henceforth kNN). kNN is a clustering technique which learns how to distinguish two or more classes on the basis of a discriminating boundary. Starting from the number of near similar occurrences (the parameter k), kNN sets the boundaries of the pattern of distribution. The dataset underwent preliminary feature scaling and was split into 75% for training with k = 40 and 25% for test. We ran the algorithm each time for each relevant feature found by means of Boruta. Given the multi-dimensional nature of the test, we coupled each vocalic feature with its respective consonantal one (e.g. VarcoV coupled with VarcoC). The two-dimensional algorithm (and its plot) allows us to appreciate further the fact that only one dimension has more predictive power than the other. For example, if the boundary line of the kNN plot is orthogonal to only dimension, the other feature will be proven as insignificant in the clustering, confirming the output of Boruta. Furthermore, by drawing boundaries, kNN also expresses the direction of each dimension as a grouping factor, showing for example that values falling within a certain range will probably result in a specific cluster.

5.2.2 VarcoV

The least effective feature among the selected three is VarcoV. Compared to %V and VCCI, the vocalic standard deviations are therefore less efficient in distinguishing between VD and VI.

The results of the kNN training for VarcoV are appreciable in Fig. 4. The areas and dots in red represent VD and the ones in blue represent VI. The lightly colored dots on the background represent the probable class that a point would belong to if it were there. Furthermore, the dot radio represents the likelihood of the classification. For example, the radio is smaller for those dots near the boundary, since the classification is uncertain there. As more extreme values are more likely to belong to a specific group, the class likelihood displays a bigger radio. Fig. 4 shows that the only discriminant line between varieties is orthogonal to VarcoV. VD utterances in the corpus have lower VarcoV in opposition to VI, which occupies the higher area on the x-axis. As expected, VarcoC does not seem to play a discriminant role in the distribution, since the scattered dots occupy the same area on the y-axis. Although two areas can be detected, there is confusion near the boundary, indicating that some VI sentences can appear in the red area and vice versa. Such confusion also leads to a scarce accuracy of 50% in the test set, which is not satisfying for the classification task.

To check whether the language shift within each speaker corresponds to a change in the durational-rhythmic properties of the utterance, we report in Fig. 5 the boxplots for each speaker in a multi-facet grid (speakers are indicated with different numbers). As already suggested by kNN, speakers tend to have higher VarcoV when reading sentences in VI. While this tendency is homogeneous across speakers (see e.g., the median lines), the error bars of VD and VI occupy similar areas, making it difficult to draw clear-cut boundaries. To conclude, while we can see higher VarcoV values for Italian indicating a trend in the data, there is still a fuzzy zone, confirming VarcoV as a non-ideally reliable discriminant dimension.

Figure 4 - *Scatterplot of the sentences along the dimensions of VarcoV with the decision boundary and likelihood area provided by kNN*

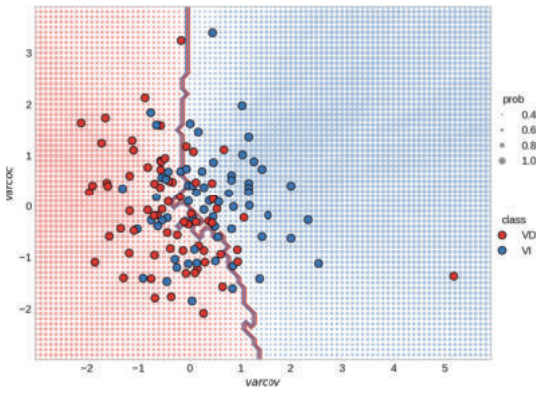
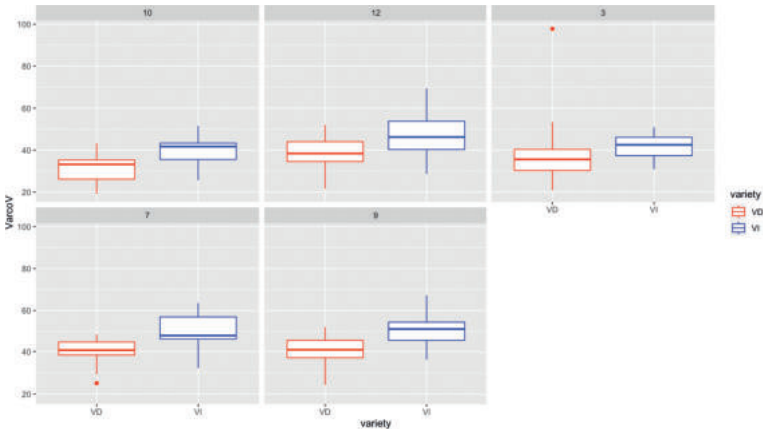


Figure 5 - *Boxplots divided by speaker indicating the range of VarcoV for both VD and VI*

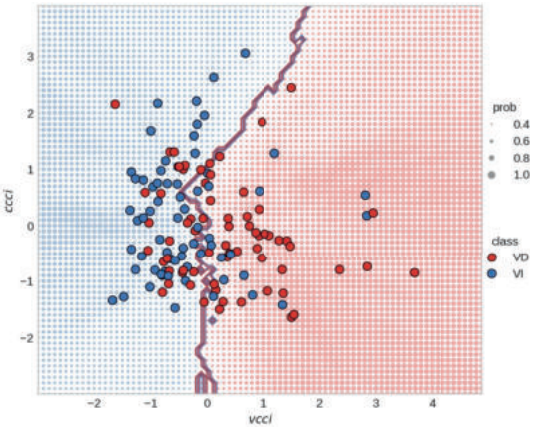


5.2.3 VCCI

We can now turn to kNN for the Compensation and Control Indexes: this time the model scored an accuracy level of 86%, which we consider as valid for discriminating the varieties in our data. Interestingly, kNN shows that VD has higher VCCI levels,

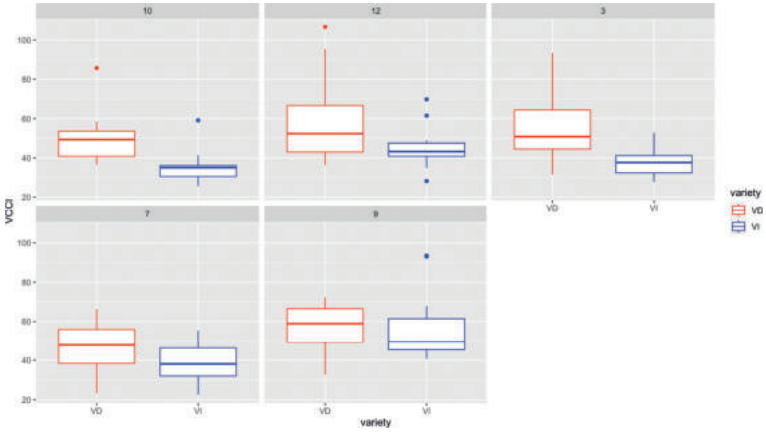
while VI occupies lower areas. Here again, the boundary is orthogonal to VCCI, while the distribution of data does not seem to be sensitive to CCCI, as expected based on the results of the feature selection.

Figure 6 - *Scatterplot of the sentences along the dimensions of VCCI and CCCI with the decision boundary and likelihood area provided by kNN: training set*



Also in this case, however, there is much confusion around the boundary area (Fig. 6). Let us explore the distribution of VCCI per speaker (Fig. 7). The tendency described by kNN corresponds to the homogeneous behavior of each speaker: the median values of VCCI are higher for VD. Note also that while some speakers make a clear-cut distinction (e.g., speakers 3 and 10), others have overlapping areas. This shows that the tendency is fairly constant, but the differentiation rate within speakers is subject to interindividual variation.

Figure 7 - *Boxplots divided by speaker indicating the range of VCCI for both VD and VI*



5.2.4 %V

Finally, we can turn to the most indicative predictor in distinguishing between VI and VD. We employed the same methodology as the previous dimensions. The kNN predictor reached an accuracy level of 92% for the test set, confirming that this is the most reliable dimension. Note that *Correlatore* does not have an inborn function to calculate %C, the opposite dimension of %V, so we calculated with the formula %C = 100 - %V. This explains why in the kNN plot the two dimensions are linearly dependent. Naturally, the decision boundary is a diagonal (Fig. 8): %V and %C are strictly dependent, and the decision boundary is drawn along both dimensions. In particular, VD sentences are more likely to appear with higher values of %V, while VI correlates with higher %C. This tendency seems to be well maintained by each speaker (Fig. 9).

Figure 8 - *Scatterplot of the sentences along the dimensions of %V and %C with the decision boundary and likelihood area provided by kNN*

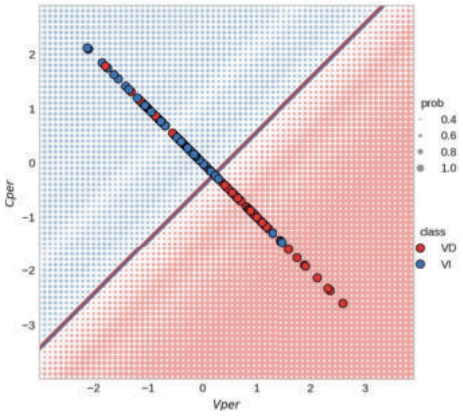
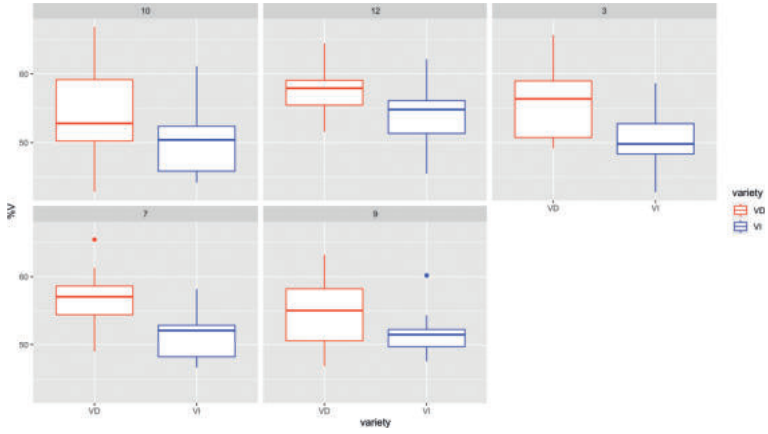


Figure 9 - *Boxplots divided by speaker indicating the range of V% for both VD and VI*



5.3 Intermediate discussion: from timing to rhythm

To the goal of comparing dialect and Italian from the prosodic point of view, we tested a number of rhythm metrics and found measurements capturing possible durational differences between VD and VI within each speaker. The results show that three of these measurements highlight consistent differences in the temporal properties when the speakers switch language. Even if the speakers adapt their production when speaking VD or VI to different extents, the overall tendency seems to be commonly shared.

The analysis also revealed that metrics perform very differently in distinguishing between the VD and VI sample. VD and VI can be distinguished only on the basis of a subset of the metrics tested in this study, namely %V, VCCI and, to a limited extent, VarcoV. It is noteworthy that vocalic and consonantal measurements do not work in tandem: the most important dimensions all rely on vocalic intervals, where the consonantal intervals seem less important in discriminating between VI and VD.

Overall, the analysis of durations of vocalic and consonantal intervals in our datasets indicates differences between VI and VD as far as vowel length is concerned, in line with results in the literature pointing to the relevance of vowels in characterizing VD (Schmid, 2014). As discussed in § 3, however, durational measurements taken as such do not provide relevant information about the rhythmic organization of a language.

Interestingly, the most performant feature in the identification of the variety is %V, which is the most straightforward metric simply displaying the amount of vocalic quantity in the sentences. This linear segmental dimension suggests that VD either has more vocalic intervals than VI, or that vocalic intervals are longer, or both. Similarly, VD displays higher values compared to VI also for VCCI, indicating a higher vocalic ratio in the dialect also when the number of phonological segments composing the interval is considered. Although caution is needed when interpreting the results from these metrics (Arvaniti, 2009, Prieto et al., 2012), we take them at face value because the segmental material was strictly controlled (cf. Fig. 3). Since the lexical material and the consonant-vowel ratio is kept constant across varieties, then this outcome might be attributed to either vocalization of (expected) consonantal segments or lengthening of vocalic intervals, or to a combination of the two. The datasets annotation shows that only a small amount of /l/ in the Venetian is realized in articles and prepositions. In most of the cases (97%), /l/ is not realized and only [a] is present (cf. Fig. 2), also when no palatal vowel is present in the segmental environment, contra our expectations based on the literature (Tomasin, 2010). Although the lack of the consonant /l/ has arguably an impact the consonant-vowel ratio in VD *vis-à-vis* VI, as reflected in the results provided by %V and VCCI, the measurements discussed in the preceding section do not provide information about the actual realization of /l/. The presence of [a] alone hints indeed to two possible scenarios: in the first, /l/ in the article *ła* "the" is simply deleted. In this case, we can expect the duration of the vowel to be unaffected and be therefore comparable to that of the other unstressed vowels of the utterance. In the alternative scenario, /l/ is

realized as a part of [a], which would consequently be lengthened. This hypothesis is coherent with the picture given by the metrics, since both the percentage of vocalic intervals, and the V-C ratio *vis-à-vis* the phonological segments would be higher in VD than in VI. We will verify this hypothesis in the following section (5.3.1).

While ΔC was not a discriminant factor, ΔV performed better but not at a satisfactory level to consider it essential in the distinction between VI and VD (cf. Fig. 4), while the normalized version of ΔV, VarcoV, reached a higher level of importance. VD correlates with lower VarcoV values, indicating that vowels in VI might have a more variable duration than in VD. This parameter, however, had a low predictive power compared to %V and VCCI. Altogether, the metrics suggest that vowels in VD might be more abundant, or longer, or both, and less variable in length compared to VI. It is not straightforward to elaborate about a scenario that can account for all these features. Along the lines of Alfano, Savy and Llisterri (2009), however, it may be hypothesized that the observed durational properties result (at least partially) from different durational strategies in marking stress. In what follows, we will also explore this point.

5.3.1 Vowel length in VD vs. VI

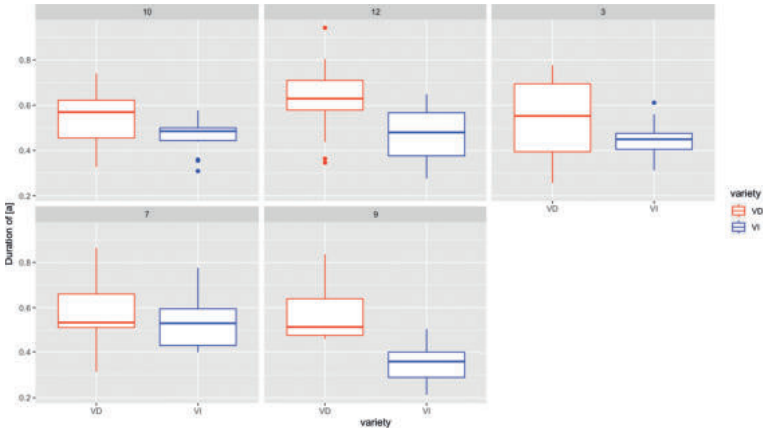
We compared the length of [a] in the article *ła* (VD) and *la* (Italian) "the" to that of the other unstressed vowels of the utterance, to verify if [a] in VD is lengthened. Given the relatively little sample dimension, we cannot test frequentist hypotheses to make inferences on a population level. However, to make sense of the data while controlling speaker variation and the effect of predictors other than the variety, we performed a Bayesian linear mixed model (McNelsh, 2016) to predict the length of the vowel [a] with the variety and sentence type. The model also included the speaker and item in the random structure. Since we expect variation in the segmental length given by the speech rate of each speaker and utterance, we decided to center this dimension. To do so, we followed the procedure displayed in the formula. We first calculated the mean of all unstressed vowels (n) in the sentence. In the calculation of the mean (l), we factored out the actual interval of interest (duration of /a/) by subtracting it. Then, the length of [a] for each utterance was respectively divided by the utterance mean.

(1) $$Centr\Delta t_i = \frac{\Delta t_i}{l} \quad where\ l = \frac{\sum_{j=1}^{n} \Delta t_j - \Delta t_i}{n-1}$$

Back to the statistical modeling, we expect a posterior probability of at least 95% to conclude that these data confirm the hypothesis that [a] is lengthened. The model was built in Stan using the R interface *brms* (Bürkner, 2017), using a Monte Carlo Markov Chain with 4 chains and 2000 observations for each chain. A weakly-informative prior was specified, given the exploratory nature of study (hyperparameters set at a normal distribution with mean = 0 and sd = 1).

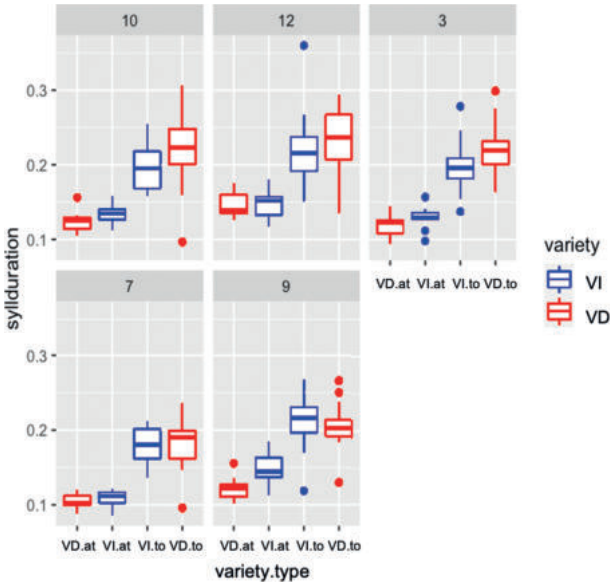
The model, which obtained a R-hat value of 1, specifies an intercept set at the VI with the sentence type BF at 0.46, est. error = 0.03, and a credible interval [0.40, 0.52]. While the effect of the sentence type did not yield any credible difference on the dataset (Category CF CI[-0.06, 0.05], Category NF CI[-0.01, 0.09]). The effect of variety produced a strong effect: the level VD yielded an estimate of 0.12, std. error = 0.02, and CI[0.07, 0.16], probability = 95%). We can conclude that the data and model support the hypothesis that [a] in VD *la* has longer duration than in VI. Moreover, the pragmatic category (e.g. whether the sentence was Broad, Contrastive, or Narrow focus) did not yield any credible effect in the distribution of the length. The differentiation between varieties seems also stable across the speakers (Fig. 10). These results support the hypothesis that /l/ is not simply canceled, but triggers a durational readjustment, i.e., the lengthening of [a].

Figure 10 - *Boxplots divided by speaker indicating the duration of [a] in VD and VI*



To further explore the durational properties of VD and VI and try to trace back the sources of the differences highlighted by the metrics, we compare the duration of stressed and unstressed syllables in VD and VI (Fig. 11). In doing so, we excluded [a] from the dataset as it would obviously have an impact on the results concerning unstressed syllables. Although realizations vary to a certain extent (see speaker 9), unstressed syllables tend to be shorter or similar in length to those of VI, while stressed syllables tend to be longer in VD. This result indicates that speakers can indeed adopt different temporal strategies to mark stress when switching from one language to the other.

Figure 11- *Boxplots divided by speaker indicating the duration of stressed ("to") and unstressed ("at") syllables in VD and VI*



6. *Discussion*

In this section we try to link the phonetic measurements described so far to the higher-level rhythmic organization, i.e., to the prominence hierarchy and its implementation.

The exploration of durational features highlighted several differences between VI and VD. Firstly, the analysis showed that vocalic intervals are globally longer in VD than in VI. Although this result is coherent with the well-known lack of consonantal gemination in VD (Zamboni 1988), it is worth noting that the Italian target sentences used in the present study did not include geminated consonants and, therefore, no role can be attributed to gemination in the results obtained on this dataset. Lack of gemination and longer vocalic intervals, however, fit the picture of VD and VI with divergent temporal organizations at the segmental level. Note that, differently from VD, in VI the distinction between singleton and geminate consonants is consistently made (Mairano, De Iacovo, this volume; in line with what observed for other varieties of Italian spoken in the same area; Zmarich, Gili Fivela, 2005).

Secondly, duration seems a more relevant cue to mark stress in VD than in VI, since stressed syllables tend to be longer in the dialect, and therefore more pronouncedly distinguished from unstressed syllables in this variety. As pointed out by Alfano, Savy and Llisterri (2009), related languages which are structurally similar, such as Italian and Spanish, may differ in their temporal organization and in their use of duration as a cue to stress in production and perception. Our data go in the

same direction, showing that differences concerning the role of duration in marking stress can emerge also from the comparison of an Italo-Romance dialect and the corresponding Italian variety. Additionally, the data indicates that vowel duration in VD might be less variable than in VI. This result can in principle be in line with the scenario of a clearer durational distinction between stressed and unstressed vowels in VD. However, this link is speculative, inasmuch the relationship between the VarcoV values and the stressed/unstressed distinction has not yet been investigated. We leave this topic open for future research.

The results presented in this work also add a piece in the description of the so-called *'l' evanescente*, showing that consonantal deletion can take place also when there is no palatal vowel in the segmental environment, differently from what has been reported in the literature so far (Tomasin, 2010). These results, however, also raise the question about why the vowel [a] in the article is lengthened. A possible explanation is that lengthening is necessary in order maintain the perceptual salience of the article, which would be otherwise too short to be identified in the speech flow. This explanation is in line with the idea that durational adjustments can be driven by linguistic factors above the segmental level.

A further observation for future work concerns the possibility that /l/ is not the only consonant subject to weakening in VD, besides the well-known lack of geminate consonants. The example provided in Fig. 2 suggests that weakening may also involve rhotics. If this is confirmed, consonantal weakening in VD might regard a larger set of sonorants.

Altogether, these results indicate differences in syllable realization (consonantal weakening, vowel lengthening), stress realization (different role of duration in marking stress). The observed differences in timing, therefore, are likely to influence the implementation of the prosodic hierarchy in VD and VI. Consistently, [a] lengthening is likely to be driven by linguistic forces above the level of the segment, as this adjustment could be due to the necessity of preserving the article's perceptual salience at the level of the phonological words.

At present, little work has been done on the intonational phonology of Italo-Romance dialects, including VD. Preliminary research on Veneto dialects (Magistro, Crocco, 2022) suggests that broad focus statements in these varieties might be more dynamic than Italian in the realization of edge tones. On the other hand, the boosting of duration in stressed syllables might also be part of a strategy to enhance prominence, as proposed for Neapolitan dialect by Crocco, Gili Fivela and D'Imperio (2022). Therefore, future research needs to delve into the link between longer duration in stressed syllables and the intonational organization of the dialect, to pinpoint possible influences going from the tune to the text, or the other way round (Grice et al. 2018, Roettger, Grice 2019).

7. *Conclusions*

In this paper we outlined a methodological pipeline based on the use of machine learning techniques to assess the power of durational measurement to distinguish between two languages, with the goal of linking phonetic variability in segmental duration to the higher levels of the prosodic organization.

Assuming that different languages are likely to differ also in their prosodic properties, and hence also in their temporal and rhythmic organization, we tested the hypothesis that bilingual speakers can adapt their temporal organization of when speaking one or the other language. To test this hypothesis, we examined the production of a group of bilingual speakers, asking them to read aloud a set of controlled sentences in both the languages we wanted to investigate (in our case: VI and VD). The material was then carefully segmented to extract duration measurements of consonantal and vocalic intervals. Using a set of machine learning techniques, we subsequently identified the durational parameters that are more likely to distinguish the languages under investigation from the durational point of view. Finally, we linked the results provided by the durational analysis to what is known about the phonology of the examined languages, providing a basis for future investigations aimed at understanding the role of durational facts in the prosodic organization of a language.

We assumed a layered organization of linguistic rhythm, but intentionally no claim was made concerning rhythmic typology and the possibility to assign a language to one or the other rhythm class. Instead, we further convolute rhythmic properties by combining segmental processes and, possibly, salience factors by pointing at the multi-faceted nature of rhythm.

We tested the so-called rhythm metric focusing on the measurements they are based on and exploring the predictive power of such measurements separately through statistical analysis. Machine learning algorithms proved to be a useful tool to assess the possibility that two languages can be distinguished based on a given feature. Our preliminary results indicate that by keeping other linguistic factors controlled and focusing on duration at the segmental level, we can indeed highlight divergences in the temporal organization of languages, which are likely to be, though indirectly, relevant for the rhythmic structure. Although further research is needed to bridge the gap between phonetics and phonology, fine-grained analyses such as those presented in this article can represent a viable methodological option to proceed on this path.

Of course, duration does not exhaust the phonetic features that can be potentially relevant in the rhythmic differentiation between languages or varieties. Other measurements can indeed be performed that could be complementary and perhaps equally useful and need therefore to be integrated. The idea behind this paper, to be further investigated, is that timing effects and rhythm may be dependent from the high-level rhythmic organization. Since languages rely on duration to a different extent to realize stress, other prosodic parameters need to be included in the picture to link realization and function.

References

Alfano, I., Savy, R. & Llisterri, J. (2009). Sulla realtà acustica dell'accento lessicale in italiano ed in spagnolo: La durata vocalica in produzione e percezione. In Romito, L., Galatà, V. & Lio, R. (eds.), *La fonetica sperimentale. Metodo e applicazioni*. *Atti del IV convegno nazionale AISV* – Arcavacata di Rende (CS). [CD-ROM]. Torriana: EDK Editore. 22-39.

Andreeva, B., Barry, W. & Koreman, J. (2014). A Cross-language Corpus for Studying the Phonetics and Phonology of Prominence. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14), pages 326–330, Reykjavik, Iceland. European Language Resources Association (ELRA).

Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. In *Journal of Phonetics,* 22, 239-268. https://doi.org/10.1016/S0095-4470(19)30203-7

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. In *Phonetica* 66, 46-63. doi: 10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. In *Journal of Phonetics,* 40, 351-373. https://doi.org/10.1016/j.wocn.2012.02.003

Barbosa, P. (2002). Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. In *Cadernos de estudos linguisticos*, *43*, 71-92. https://doi.org/10.20396/cel.v43i0.8637150.

Barbosa, P. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. In *Speech Communication*, 49, 725–742. https://doi.org/10.1016/j.specom.2007.04.013

Berruto, G. (2012). *Sociolinguistica dell'italiano contemporaneo* (2nd ed.) Roma: Carocci.

Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici.* Firenze: Accademia della Crusca.

Bertinetto, P.M., Bertini, C. (2008). On modeling the rhythm of natural languages. In *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brasil, 427-430.

Boersma, P., Weenink, D. (2022). Praat: doing phonetics by computer [software]. Version 6.2.09, http://www.praat.org/

Bürkner, P.C. (2017). brms: An R package for Bayesian multilevel models using Stan. In *Journal of statistical software*, 80, 1-28.

Clarke, E.F. (1999). Rhythm and timing in music. In Deutsch, D. (Eds.) *The psychology of music*, New York: Academic press, 473–500.

Cortelazzo, M., Paccagnella, I. (1992). Il Veneto. In Bruni, F. (Eds.) *L'italiano nelle regioni*, Torino: UTET, vol. I, 263-310.

Coseriu, E. (1981). Los conceptos de «dialecto», «nivel» y «estilo de lengua» y el sentido propio de la dialectología. In *Lingüística Española Actual*, 3, 1–32.

Crocco, C., Gili Fivela B. & D'Imperio, M. (2022). Comparing prosody of Italian varieties and dialects: data from Neapolitan. In *Proceedings of the 11th International Conference on Speech Prosody*, Lisbon: Portugal, 140-144.

Dal Negro, S. Vietti, A. (2011). Italian and Italo-Romance dialects. In *International Journal of Sociology of Language*, 210, 71- 92. https://doi.org/10.1515/ijsl.2011.031

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. In *Journal of Phonetics, 11*(1), 51–62.

Dauer, R.M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences,* Tillinn, Estonia, 447-4450.

Dellwo, V., Wagner, P. (2003). Relations between language rhythm and speech rate. In S, V., Wagner Solé, P., Recasens, D., Romero (eds.), *Proceedings of 15th International Congress of Phonetic Science*s, 471-474.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for ΔC. In *Proceedings of the 38th Linguistic Colloquium*, Budapest, 231-241.

D'Imperio, M., Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. In *Phonology*, 16, 1-28.

Falk, S., Rathcke, T. & Dalla Bella, S. (2014). When speech sounds like music. In *Journal of Experimental Psychology*, 40(4), 1491-1506. http://dx.doi.org/10.1037/a0036858

Ferguson, R. (2017). A Linguistic History of Venice. Firenze: L.S. Olschki. http://digital.casalini.it/9788822256454

Dellwo, V., Wagner, P. (2003). "Relations between language rhythm and speech rate," in *Proceedings of the 15th International Congress of Phonetic Sciences*, 471–474.

Fletcher, J. (2010). The prosody of speech: timing and rhythm. In Laver, J., Gibbon, E.F. (eds.), *The Handbook of Phonetic Sciences* (2nd ed.), West Sussex, UK: Blackwell, 521–602. https://doi.org/10.1002/9781444317251.ch15

Frota S., Moraes, J.A. (2016). Intonation of European and Brazilian Portuguese. In Wetzels, L., Costa, J., Menuzzi, S. (Eds.), *The Handbook of Portuguese Linguistics*. Willey-Blackwell. https://doi.org/10.1002/9781118791844.ch9

Gamer, M., Lemon, J., Fellows, I., Singh, P. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. [R package] version 0.84.1. https://CRAN.R-project.org/package=irr

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M., Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian, in Frota, S., Prieto, P. (Eds.) *Romance intonation*, Oxford: Oxford University Press, pp. 140–197.

Grabe, E., Low, E.L. (2002). Acoustic correlates of rhythm class, in *Laboratory Phonology*, 7, 515–546.

Grice, M., Savino, M. & Roettger T. (2018). Word final schwa is driven by intonation: The case of Bari Italian. In *JASA*, 143(4), 2474-2486. https://doi.org/10.1121/1.5030923

ISTAT. (2017). https://www.istat.it/it/archivio/207961, last access: 30 August 2022.

Jun, S.A. (2012). Prosodic Typology Revisited: Adding Macro-rhythm. In *Proceedings of 6th International Conference on Speech Prosody*.

Jun, S.A. (2014). Prosodic Typology: By Prominence Type, Word Prosody, and Macro-rhythm. In Jun, S.A. (eds.) *Prosodic Typology* II. Oxford: OUP, 520-539. https://doi.org/10.1093/acprof:oso/9780199567300.003.0017

Ladd, R.D. (2008). *Intonational phonology* (2nd ed.) Cambridge: CUP. https://doi.org/10.1017/CBO9780511808814

Kohler, K. (2009), Whither Speech Rhythm Research? In *Phonetica*, 66, 5–14.

Kramer, M. (2009), The phonology of Italian, Oxford: Oxford University Press.

Kursa, M.B, Rudnicki, W.R. (2010). Feature selection with the Boruta package. In *Journal of Statistical Software*, 36, 1-13.

Lepschy, G.C. (1962). Fonematica veneziana, in *L'Italia dialettale,* 25 (1962), pp. 1-22.

Magistro G., Crocco, C. (2022). Rising and rising-falling declaratives in Veneto dialects In *Proceedings of the 11th International Conference on Speech Prosody*, 175-179.

Mairano, P., Romano, A. (2010) Un confronto tra diverse metriche ritmiche usando Correlatore. In: Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.) *La dimensione temporale del parlato*, Proceedings of the V National AISV Congress, University of Zurich, Collegiengebaude, 4-6 February 2009, Torriana (RN): EDK, 79-100.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. In *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5), 750-773. https://doi.org/10.1080/10705511.2016.1186549

Ramus, F., Nespor, M. & Mehler, J. (1999). *Correlates of linguistic rhythm in the speech signal*, In *Cognition*, 73(3), 265–292. https://doi.org/10.1016/S0010-0277(00)00101-3

Rathcke, T., Smith, R.H. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. In *JASA,* 137: 28-34. http://doi.org/10.1121/1.4919322

Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. In *Speech Communication*, 54(3), 368-383. https://doi.org/10.1016/j.specom.2011.10.001

Roettger, T.B., Grice, M. (2019). The tune drives the text – Competing information channels of speech shape phonological systems. In Good, J., Greenhill, S. (Eds.) *Language Dynamics and Change*, 9(2), 265-298. doi: https://doi.org/10.1163/22105832-00902006

Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti: fonetica*. Torino: Einaudi.

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. In *Proceedings of the International Congress of Phonetic Sciences*. San Francisco, August, 607-610.

Schmid, S. (1998). Tipi sillabici nei dialetti dell'Italia settentrionale. In: Ruffino, G. (Eds.) *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*. Tübingen: De Gruyter, 613-625.

Schmid, S. (2014). Syllable typology and the rhythm class hypothesis: Evidence from Italo-Romance dialects. In Caro Reina, J., Szczepaniak, R. (Eds.) *Syllable and Word Languages*. Berlin: Mouton de Gruyter, 421-454. https://doi.org/10.1515/9783110346992.421

Serianni, L., Trifone, P. (1993). *Storia della lingua italiana*, Torino: Einaudi.

Tomasin L. (2010). La cosiddetta elle evanescente del veneziano: fra dialettologia e storia linguistica". In Ruffino, G. and D'Agostino, M. (Eds.) *Storia della lingua italiana e dialettologia*. Palermo: Centro di studi filologici e linguistici siciliani. 729-51.

Turk, A., Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. In *Laboratory Phonology*, 4(1), 93-118. https://doi.org/10.1515/lp-2013-0005

White, L., Payne, E. & Mattys, S. (2009). Rhythmic and prosodic contrast in Venetian and Sicilian Italian. In Vigário, M., Frota, S. & Freitas, M.J. (Eds.) *Phonetics and Phonology: Interactions and interrelations*, Amsterdam: John Benjamins. pp. 137–158. https://doi.org/10.1075/cilt.306.07whi

Zamboni, A. (1988). Veneto, in Holtus, G. Metzelin & M. Schmitt, C. (Eds.) *Lexicon der romanistischen Linguistik*, Tübingen: Niemeyer, 517-38.

Zmarich, C., Gili Fivela, B. (2005). Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e labiodentale. In *Misura dei parametri, Atti del Convegno Nazionale AISV*, Padova, 2-4 dicembre 2004, Torriana (RN): EDK, pp. 429-448.