On the timing of V-to-V intervals in Italian: a review, and some new hypotheses

Chiara CELATA[1] & Paolo MAIRANO[2]
[1]Scuola Normale Superiore di Pisa (Italy)
[2]Nuance Communications Inc. (Turin, Italy)
c.celata@sns.it, paolo.mairano@nuance.com

ABSTRACT
This paper is intended to discuss aspects of durational variability of V-to-V intervals in Italian from a phonetically based perspective. Recently, the interest in the phonetic properties underlying rhythm typologies has been growing, also as a consequence of the repeatedly observed weakness of phonological factors in the explanation of rhythmic differences. Timing is only one of the many aspects contributing to language rhythm. The speaker and the text are fundamental sources of variation, in the sense that rhythm depends far more on individual timing strategies than on the phonological structure of a language.

Keywords: Italian phonetics, intervals, rhythm.

[Recibido, septiembre 2013; aprobado, diciembre 2013]

Cronometrando los intervalos *V-to-V* en italiano:
revisión y nuevas hipótesis

RESUMEN
Este artículo intenta describir aspectos de la variabilidad duracional de los intervalos V-to-V en italiano desde un punto de vista fonético. Recientemente ha aumentado el interés en las propiedades fonéticas subyacentes en las tipologías rítmicas, también como consecuencia de una debilidad de los factores fonológicos, observada repetidamente la explicación de diferencias rítmicas. El hablante y el texto son fuentes fundamentales de variación, en el sentido de que el ritmo depende más de estrategias temporales individuales que en la estructura fonológica de la lengua.

Palabras clave: fonética italiana, intervalos, ritmo.

1. Introduction

A longstanding tradition of studies in speech rhythm and timing has focused on the importance of beats or voice onsets. Voice onsets are perceptually salient and require precise articulatory coordination of glottal and supraglottal structures. They have been shown to function as P-centers in the speech chain, contributing greatly to the temporal organization of speech especially in oscillator-based models of rhythm (e.g. Barbosa 2006; Bertinetto & Bertini 2008). In moraic languages such as Japanese, the interval between the onset of two vocalic nuclei in CVC:V disyllables is twice as long as the interval in CVCV disyllables, and functions as a mora timing unit across different speech rates (Brady et al. 2006; Hirata & Forbes 2007). In Standard Austrian German, which shares some rhythmic properties with «syllable-based» quantifying languages and some others with «word» languages (Moosmüller & Brandstätter 2014 in press), the V-to-V interval is longest in CV:C:V, shortest in CV:CV and of an intermediate length in CVC:V disyllables; in this language, C: and C are traditionally referred to as «fortis» and «lenis» consonants, respectively, but phonetically they are mostly distinguished for duration (Moosmüller & Ringen 2004).

This paper is intended to discuss aspects of durational variability of V-to-V intervals in Italian from a phonetically based perspective. Recently, the interest in the phonetic properties underlying rhythm typologies has been growing, also as a consequence of the repeatedly observed weakness of phonological factors in the explanation of rhythmic differences (see Arvaniti 2012 for a review). Timing is only one of the many aspects contributing to language rhythm (an important role being played also by spectral properties, e.g. Tilsen & Arvaniti 2013, and changes in F0, e.g. Barry et al. 2009). In the domain of timing, recent studies show that the correspondence between phonological properties such as syllable structure and acoustic measures is much weaker than previously expected. The speaker and the text are fundamental sources of variation, in the sense that rhythm depends far more on individual timing strategies than on the phonological structure of a language (e.g. Van Santen & Shih 2000; Dellwo et al. 2007; Loukina et al. 2013). Low-level phonetic effects (including text composition, Dellwo et al. 2007, and speech rate, e.g. Bertini et al. 2011) probably play a greater role than initially thought. This paper moves from similar considerations and presents a critical review of the problem of V durational variability as a function of syllable structure in Italian. Previous studies on the subject are reviewed (§2), then some new data from a small production experiment on Italian voice onset timing are presented (§3 and §4), suggesting that the role of the phonological syllable in determining V durations should be put in the right perspective. However, the experiment is still in progress and we present here only the preliminary results from a small subset of data; as a consequence, the contribution of this paper is limited to opening new, hopefully fruitful research perspectives, without any presumption of providing irrefutable evidence against the phonological syllable as a timing unit in Italian. In the conclusion of the paper (§5), we discuss the reasons why, in our opinion, the issue of V-to-V intervals in Italian is worthy of further investigation, and propose some future directions of study.

2. V-to-V sequences in Italian: a review and some new hypotheses

In earlier studies about rhythm (e.g., Pike 1945; Abercrombie 1967), languages were classified as either stress-timed or syllable-timed. According to those approaches, the fundamental rhythmical unit in Italian was said to be the phonological syllable and

Italian was classified as a syllable-timed language with isochronous syllables and was opposed to stress-timed languages such as English and German, which were said to be characterized by isochronous feet (inter-stress intervals). However, as is well-known, no studies ever confirmed these claims experimentally. For instance Bertinetto (1977; expanded in 1981 and 1983) found that the durational increase of the syllable tends to be a function of the number of phones that compose it.

In contrast to the traditional view of Italian rhythm as based on the phonological syllable, Edda Farnetani and Shiro Kori argued in a 1986 paper that the basic temporal unit in Italian was the *rhythmical* syllable, that is, «the temporal interval from vowel onset to vowel onset» (Farnetani & Kori 1986: 17). In that paper, a set of 1029 disyllabic paroxytone test words (plus 651 trisyllables and oxytone disyllables) was analyzed; the duration of syllable onsets (varying from 0 to 3 consonants), offsets (varying from 0 to 1 consonant) and nuclei (/a/ monophtongs) was acoustically measured. The word onset consonants were singletons /s k g l r n/ and the clusters were /sk st skr str zgr tr gr/. In word-internal position, singletons /s l r n t d/, geminates /sː lː nː tː dː/, palatals[1] /ʃː ʎː ɲː/, heterosyllabic clusters /ld lt nd nt rt rd st/, tautosyllabic clusters /tr dr/ and triconsonantal clusters /ltr str ntr/ were included. The most robust result of the durational analyses concerned the relationship between temporal variations and rightward syllable extension: a significant reduction of vowel durations in closed syllables was observed, in both cases of heterosyllabic clusters and geminates. That result was consistent with previous cross-linguistic data supporting the conclusion that the acoustic duration of a stressed vowel decreases as the syllable complexity increases (e.g. Lindblom & Rapp 1973; Fowler 1977). For Italian, vowel shortening in closed syllable has been repeatedly confirmed in successive experimental studies (e.g. Marotta 1985; Pickett et al. 1999; Turchi & Bertinetto 2000 and the bibliography therein), although some scholars have stressed that durational oscillations are statistically significant only under intonational phrase stress, while they tend to disappear in connected speech (e.g. Bertinetto 1981; Marotta 1984).[2] In the study by Farnetani & Kori, tri-consonantal clusters also seemed to induce vocalic reduction to the same extent (although they provided less reliable indications because of their scarcity in the experimental corpus). On the contrary, leftward syllable extension (i.e., increasing the number of syllable- and word-initial consonants) did not produce significant compression effects on stressed vowels (with the exception of /gr tr/ and /zgr str/, which caused vocalic shortening as opposed to /r/).[3]

Overall, the findings of the Farnetani & Kori's study suggested that vowels were not (or not strongly) affected by the presence of an onset consonant nor by the complexity of the syllable onset, while they did undergo systematic durational variations when changes occurred in the following context. Farnetani & Kori concluded that vocalic nuclei are overall more cohesive with the syllable offset than with the syllable onset.
Besides the strong influence of coda consonants on vowel durations, other anticipatory shortening effects are reported on in the Farnetani & Kori (1986) study. These appeared to be phonetically based and unrelated to syllable structure.

---

[1] In agreement with the terminological conventions of Farnetani & Kori, we here use the term "palatals" to refer to both palatoalveolar /ʃː/ and true palatals /ʎː ɲː/.
[2] Word length has also been shown to influence vowel duration: stressed vowels undergo a regular compression effect as post-tonic syllables are added, although such effect is not as strong as in English (D'Imperio & Rosenthall 1999; Vayra et al. 1999; Hajek et al. 2007).
[3] For this aspect Italian is said to differ from other languages, such as English (see Hermes 2013 and bibliographical references therein).

One such effect was triggered by the voicing of the onset of the second syllable which turned out to influence the duration of the vowel of the first syllable in test words such as /ˈlanda/ vs. /ˈlanta/ and /ˈlardo/ vs. /ˈlarto/: clusters containing a voiceless obstruent tended to be longer than their voiced counterparts, and produced a significant reduction of the preceding vowel in the speech of at least some of the subjects.[4] The lengthening-before-voicing effect (which has been observed for many languages, particularly for single codas) has been confirmed to be robust in Italian read speech for V1-nasal-stop-V2 sequences across different speech rates (normal and fast), with voiceless stops producing shorter durations of both the preceding nasals and the preceding stressed nuclei (V1), compared to voiced stops (Celata & Calamai 2011; see also Van Santen et al. 1992 for similar results on American English tautosyllabic sonorant-obstruent clusters).

A second type of anticipatory shortening in the Farnetani & Kori's data was the influence of the extension of the following syllable onset on the duration of the preceding syllable stressed vowel. According to that study, the vowels of /ˈladra/ and /ˈlatra/ were significantly shorter than in /ˈlada/ and /ˈlata/, respectively, though remaining longer than in /ˈladːa/ and /ˈlatːa/. The Authors also found a negative correlation between the duration of the vowel and that of the following consonant in the /ˈladra/ and /ˈlatra/ test words. They interpreted this result as if «the talker […] [was] adjusting the duration of the vowel to that of the following consonant» (p. 27). Unfortunately the study did not provide information on the duration of the /dr/ and /tr/ clusters. Consequently we do not know if the clusters had a duration somewhat intermediate between that of the singletons /d/, /t/ and that of the corresponding geminates. The hypothesis seems plausible indeed,[5] and if this was the case, we could hypothesize that an anticipatory shortening in /ˈladːa/ as opposed to /ˈladra/ and in /ˈladra/ as opposed to /ˈlada/ (and the same for the /tː/, /tr/ and /t/) was gradually implemented as a function of the duration of the following consonantal interval. In addition, this phonetically-based effect might be as strong as to overcome the categorical syllable-based effect according to which there should be no significant vocalic oscillation in /ˈladra/ as opposed to /ˈlada/.

Palatals provided a third case of evidence in favor of the existence of phonetically-based anticipatory compensation effects. In the study by Farnetani & Kori (1986), the duration of vowels before palatals /ʃː/, /ʎː/, /ɲː/ was intermediate between the duration of vowels in open syllables and that of vowels before geminates /sː/, /lː/, /nː/. The duration of palatals also tended to be intermediate between that of the singletons /s/, /l/, /n/ and that of the corresponding geminates.[6] The data proved the existence of an inverse relationship between the duration of the vowel and that of the following consonant. This relationship is the result of a low-level adjustment of the duration of the stressed nucleus as a function of the duration of the following consonantal interval.

---

[4] See Chen (1970), Port (1979) for an early observation of the lenghtening-before-voicing effect in English.

[5] This hypothesis is in fact supported by the data of the production experiment which will be presented in the following sections. In Corpus A (see below for details), the average durations of the relevant C intervals were the following: /t/ = 72 ms (s.d. 17), /tr/ = 113 ms (s.d. 16), /tː/ = 140 ms (s.d. 30); /d/ = 56 ms (s.d. 13), /dr/ = 95 ms (s.d. 17), /dː/ = 132 ms (s.d. 30).

[6] There is no opposition between singleton and geminate palatals in Italian. Palatals tend to be shorter than true geminates in intervocalic position and, differently from singletons, undergo a reduction process in post-consonantal and post-pausal contexts (Celata & Kaeppeli 2003). They do not undergo contextual lengthening by *Raddoppiamento Fonosintattico* in central Italian speech.

Taken together, the lengthening-before-voicing effect, the clusters of *muta cum liquida* and the palatals show the existence of anticipatory compensation effects that depend on the production characteristics of the segments involved, and that span the syllable boundary. Based on the observation of such trans-syllabic effects, Farnetani & Kori (1986) concluded that the anticipatory shortening of stressed vowels may occur «independently of syllable structure». For this reason, as anticipated above, they postulated that the fundamental timing unit in Italian is the rhythmical syllable, which extends from the onset of a vowel to the onset of the following one (p. 27). This view is explicitly based on the empirical observation of strong compensatory effects that alter the duration of the vocalic intervals as a function of the duration of consonantal intervals and *vice versa*.

Stressing the role of such compensatory effects in shaping the prosodic patterns of Italian is obviously not equal to say that they alone can account for all the variance of the data. Indeed, they are first of all counterbalanced by the segments' intrinsic limits of compressibility (e.g. van Santen & Shih 2000 for a recent discussion of this problem). The property of elasticity may explain e.g. why stressed vowels did not shorten proportionally to the increase of the number of consonants in the V-to-V interval (e.g. vowels in /ˈlalta/ were as short as in /ˈlaltra/) in the data by Farnetani & Kori (1986). A similar explanation is equally valid as one postulating the role of a constraint imposed by the phonological structure.

The fact that the basic rhythmic unit in Italian is likely to be the V-to-V interval (and not the phonological syllable) was also suggested by the results of Mairano (2011: 30-34), where V-to-V intervals in Italian were shown to be affected by the least variability, compared to the other 14 languages tested. In that study, some of the recently developed rhythmic metrics (deltas and PVIs, intended to measure the degree of variability in data) were calculated on V-to-V intervals in the speech from speakers of 15 languages. Results indicated that Italian had among the lowest levels of V-to-V durational variability (lower levels were only found for Greek in the case of rPVI), thus suggesting that the duration of V-to-V intervals is relatively stable in that language. In order to factor out inter-speaker variability, the experiment was replicated on a native Italian penta-lingual proficient speaker. Even in this case, V-to-V variability was found to be least in Italian than in the other 4 languages produced by the speaker.

In sum, the paper by Farnetani & Kori (1986) as well as subsequent evidence in Mairano (2011) clearly points to the existence of low-level compensatory effects co-determining segmental durations in Italian. We believe that the consequences of such findings have not been entirely understood and appropriately integrated in a wider phonetic / phonological perspective. In particular, the fact that *gradual* phonetic effects are at play, besides categorical ones, appears to have been almost completely disregarded in the mainstream of the timing studies on Italian. Only the broad picture of syllable-based distinctions has been retained; the finer distinctions generated by the phonetic properties of individual segments have never been systematically investigated, as appears evident from the fact that in the studies on vocalic oscillations, no data are ever provided for the corresponding Cs (either geminates or clusters). This is even more striking if one considers that vocalic shortening in closed syllables has been proved to be statistically significant only in very controlled speech conditions and, in any case, highly variable across subjects (see above). In spite of the observed variability, syllable-induced modifications have been substantially considered to be robust enough to explain all aspects of vocalic variation in Italian. For instance, vowel duration was taken as a predictor of syllable structure in the case of consonant sequences whose syllabification is unknown (/s/ + obstruent clusters, e.g. Marotta 1995; Turchi & Bertinetto 2000).[7]

In the following sections, we present the results of an acoustic study of voice onset intervals in two different corpora of spoken Italian, the first being a small, targeted, well-balanced, carefully and manually labelled corpus, the other focusing on quantity (being a large, generic and semi-automatically labelled corpus). In particular, the study is a first attempt at quantifying the overall weight of phonetically-based compensatory effects in a speech dataset, by analyzing to what extent the duration of stressed Vs can be predicted by the duration of the subsequent C intervals.

3. Production experiment: materials and procedures

We wanted to verify if, in a VC(C) sequence, the duration of the stressed V was correlated to the duration of the following C(C) interval. In particular, to test the hypothesis that the timing effect within the voice onset interval is independent of syllable structure, we included in the experimental design C(C) intervals with different syllabic behavior: some of them were onsets of the following syllable (i.e., the stressed V was the nucleus of an open syllable), while others pertained to both syllables, the first C being the coda of the first syllable and the second C being the onset of the second syllable (i.e., the V was the nucleus of a closed syllable).

The experiment is still in progress and we present here the preliminary results from a small data set.

Two corpora of Italian read speech were analyzed. The first corpus (Corpus A, henceforth) comprised 145 sentences with disyllabic words in a fixed, prosodically relevant position. An example of the experimental sentences is the following: *Molte donne vanno in piazza* ('Many women go to the square'), with *donne* 'women' as the target disyllable. Eight speakers (5 males, 3 females) of central and southern varieties of Italian were recorded while reading the experimental list. The age range of the speakers was 22-34 years. Some of them were pursuing a university education at the time of the recording while others hold an academic degree. Some biographical and dialectological information about the participants is reported in Table 1.

	F1	F2	F3	M1	M2	M3	M4	M5
age	24	30	22	31	28	29	33	34
education	high school degree	MA	high school degree	MA	high school degree	MA	MA	PhD
Reported regional variety of Italian (and town of origin)	Campanian (Salerno)	Lazio (Rome)	Tuscan (Pisa)	South-ern Lazio (Terracina)	Tuscan (Grosseto)	Tuscan (Follonica)	Tuscan (Pisa)	Lazio (Rome)

Table 1. Speakers of Corpus A ("F" = female, "M" = male).

[7] In the study by Turchi & Bertinetto (2000), the duration of the preceding V was found to vary to a great extent (only some of the speakers shortened the vowel before /sC/ clusters to the same extent than before any other hetero-syllabic cluster); this result was considered an evidence of the «undecidable» syllabification of /s/+obstruent clusters in Italian. More recent acoustic and articulatory evidence in Hermes (2013) confirms that /s/+obstruent clusters differ from other word-initial clusters in Italian, for the fact that the /s/ is not produced in-phase with the following vocalic nucleus.

The speakers were recorded in a sound-proof room. They were sitting and speaking in the microphone placed on the table at a comfortable distance from the mouth of the speaker. The speakers read aloud the list of sentences which was presented on a paper sheet. The order of the sentences in the list was randomized across speakers. The speakers were explicitly asked to keep speaking rate, volume and intonation as constant as possible during the reading of the whole list. The task had a duration of about 7 minutes per speaker.

A total of 1160 disyllabic words in stressed position were recorded. Of these, 1110 were included in the analysis (the other 50 being problematic for the acoustic analysis and/or segmentation). The sentences were constructed in order to warrant that the target word was consistently preceded by 2 and followed by 4 (sporadically 5) phonetic syllables. The intonational profile was also kept constant across the sentences. The target words had a CVCV or CVCCV structure, with CC standing for either a geminate consonant or a bi-consonantal cluster, and were stressed on the first vowel. Corpus A was conceived to include as much structural variation as possible in both vowel and consonant systems. The following consonants were included in medial position: singleton stops /p t k b f g/, nasals /m n/, liquids /l r/ and fricatives /s f/, plus the corresponding geminate phonemes. As for the clusters, the experimental list included 12 heterosyllabic clusters /lp lb lv rp rk rt rb rd rg rl ps ft/, 9 tautosyllabic clusters /pl pr tr kr kl bl br dr gr/ and 4 clusters with so-called undecidable syllabification /st sp zd tl/ (Bertinetto 2004; 2011). Each consonant or consonant cluster occurred in 2 to 4 different words, varying for the quality of the preceding stressed vowel, which could be a low (/a/), mid front (/e ɛ/), mid back (/o ɔ/), high front (/i/) or high back vowel (/u/). Corpus A was manually annotated; the duration values of the relevant segments were extracted by means of an *ad hoc* Praat script.

The second corpus (Corpus B henceforth) contained read speech from 3 professional speakers of «standard» Italian (2 females, 1 male). Differently from Corpus A, this corpus was not *ad-hoc* built for the purpose of the present study; in fact, it comprises phonetically rich sentences, which were originally designed as to include as many sound combinations (i.e. diphones, triphones, syllables, etc.) as possible and to be as much representative as possible of all fundamental aspects of Italian phonology.[8]The sentences were extracted (sometimes with slight modifications) from available written corpora (mostly newspapers). They were individually recorded in a professional studio with a voice coach during several days. Subsequently, recordings were automatically transcribed and segmented into phonemes, with a final manual check. The ultimate goal of the corpus was that of creating TTS voices through a procedure of concatenative synthesis. In total, Corpus B disposes of 6224, 2664 and 2625 sentences uttered by F1, M1 and F2, respectively.

In both corpora, we measured (i) the duration of each consonantal interval of the target disyllable (either singleton, geminate or bi-consonantal cluster; in the latter cases, both consonants were measured); (ii) the duration of the preceding stressed vowel. In Corpus A, we measured the duration of each experimental sentence as well. In Corpus B, we measured the duration of the preceding vowel even when it was unstressed. For Corpus A, we also performed a normalization of the data by dividing the duration of each sentence by the number of the syllables composing it (thus obtaining the average

[8] The details of the maximization of phonetic coverage in the corpus are covered by industrial secret and therefore cannot be disclosed here.

duration of the syllables in each sentence), and then dividing (i) and (ii) by the obtained average syllable duration. This normalization was made in order to account for the fact that speech rate could vary across the experimental list, some sentences being read more slowly than others. However, we are conscious of the fact that such normalization does not take into account the fact that speech rate could (and in fact did) also vary within each sentence, with different words being read at different rates. The results of the current analysis are presented in the following section.

4. Results

4.1. Corpus A

We calculated the Pearson's correlation coefficient between the normalized V duration values and the normalized C interval duration values over the entire corpus (N = 1110). By «C interval» we mean the duration of the entire consonantal slot before the onset of the successive (unstressed) vowel: i.e., the duration of C in the case of intervocalic singletons, the duration of Cː in the case of geminates, and the duration of the C1C2 interval in the case of consonantal clusters. The result was a statistically significant ($p < .001$) negative correlation with a rather low correlation coefficient: $\rho = -0.22$.

This result revealed that, although the duration of the stressed vowel tended to decrease as the duration of the C interval increased, such tendency was not as strong as to account for the majority of the data in the corpus. The value of the ρ coefficient indicated the presence of a weak correlation.

4.1.1. Analysis by speaker and by vowel

We subsequently analyzed the behavior of the individual speakers and of the individual Vs separately. The split results are summarized in Table 2. It turned out straightforwardly that the production of speakers F1 and M1 did not conform to the production of the other speakers: the duration of the stressed Vs as produced by these 3 speakers was completely unrelated to the duration of the corresponding C interval. Differently from the other participants (Table 1), F1 and M1 were speakers of a southern variety of Italian (see also §5). Therefore it is possible to hypothesize that the different regional background explains the divergent pattern. Table 2 also reveals that not all vocalic nuclei were affected by the duration of the subsequent consonant(s) to the same extent: disyllables containing front mid vowels /e ɛ/ did not show any apparent correlation between the two variables.

	F1	F2	F3	M1	M2	M3	M4	M5
Pearson's ρ	n.s.	-0.22*	-0.34**	n.s.	-0.18*	-0.24**	-0,36**	-0.32**
N	131	135	143	141	142	137	141	

	/a/	/e ɛ/	/i/	/o ɔ/	/u/
Pearson's ρ	-0.39**	n.s.	-0.25**	-0.33**	-0.24**
N	303	253	181	173	200

Table 2. Correlation coefficients (Pearson's ρ) between stressed V duration and the duration of the following consonantal interval for each subject and each vowel of the study (Corpus A). Here and in the following tables, ** indicates that the correlation is significant with $p < .01$, while * indicates that the correlation is significant with $p < .05$. The number (N) of the relevant items for each subset of data is also given.

Given these results, we recalculated the correlation coefficient for a selection of the data, namely, after having excluded F1 and M1 and (as far as the production of the other speakers was concerned) all items containing the front mid vowels /e ɛ/. Table 3 summarizes the results obtained after this filtering. When F1 and M1 were excluded from the analysis, the correlation coefficient turned out to be slightly higher (though remaining at the level of a weak correlation): $\rho = -0.29$ ($p < .001$). When the disyllables containing the front mid vowels /e ɛ/ were also excluded, the coefficient raised again: $\rho = -0.34$ ($p < .001$).

	Filtered by subjects	Filtered by vowels	Filtered by subjects & by vowels
Pearson's ρ	-0.29**	-0.29**	-0.34**
N	838	857	648

Table 3. Correlation coefficients (Pearson's ρ) between stressed V duration and the duration of the following consonantal interval after filtering out subjects F1, F2 and F3 («Filtered by subjects»), front vowels («Filtered by vowels») and both subjects and front vowels («Filtered by subjects and by vowels») (Corpus A).

4.1.2. Analysis by C type

We subsequently verified whether the type of the C interval had any effect on the hypothesized compensatory shortening of the stressed nuclei. To that aim, starting from the 648 selected items after the exclusion of F1, M1 and /e ɛ/-disyllables, we created two separate groups: the first group comprised clusters, the second group comprised singletons and geminates. When analyzed separately, the correlation coefficient for the clusters reached the value of $\rho = -0.46$ ($p < .001$), while that for the singletons and geminates reached the value of $\rho = -0.42$ ($p < .001$), thus signaling the presence of a moderate correlation between the duration of the stressed V and that of the subsequent C interval in the two subsets (see Table 4).[9]

	Clusters	Singletons and geminates
Pearson's ρ	-0.46**	-0.42**
N	197	451

Table 4. Correlation coefficients (Pearson's ρ) between stressed V duration and the duration of the following consonantal interval according to the type of the consonantal interval (Corpus A).

Furthermore, we wanted to verify whether, within the class of singletons and geminates, the phonetic properties of the C influenced the strength of the correlation effect. We therefore split the dataset into 4 subgroups, according to the 4 consonant class: fricatives, liquids, nasals and stops.

As Table 5 shows, there were visible differences in the correlation coefficient according to the type of the post-vocalic C, with the liquids showing the highest value and the fricatives the lowest one; however, in all cases the correlation was of a moderate entity. In addition, it must be noted that the number of the observations when considering the four subgroups separately was necessarily reduced. Therefore, the observed differences of the coefficient value among consonant types should not be overemphasized.

---

[9] Vowels before singletons were overall longer than vowels before geminates (Welch one-way ANOVA for non-homogeneous variances, $F(1\text{-}283{,}744) = 19{,}755$, $p < .05$). Vowels before tautosyllabic clusters were longer than vowels heterosyllabic clusters (one-way ANOVA, $F(1{,}336) = 15{,}270$, $p < .05$).

	Fricatives	Liquids	Nasals	Stops
Pearson's ρ	-0.32**	-0.44**	-0.34**	-0.30**
N	82	68	83	218

Table 5. Correlation coefficients (Pearson's ρ) between stressed V duration and the duration of the following consonantal interval according to the type of the consonant (singletons and geminates only) (Corpus A).

Concerning the group of clusters, we hypothesized that the fact of being a tautosyllabic or a heterosyllabic cluster influenced the result of the correlation analysis. According to the view that syllable structure entirely predicts V durational oscillations, the V preceding a tautosyllabic cluster should be completely unaffected by the duration of the subsequent cluster, since in this case the vowel is the nucleus of an open syllable and the cluster pertains entirely to the following syllable. According to an alternative view that syllable structure is the major predictor of vocalic oscillations but, when a consonantal coda is present, V shortening may be implemented as a compensatory effect triggered by the duration of the syllabic coda, we might expect that, in heterosyllabic clusters, the preceding V shortens as a function of C1 duration more reliably than as a function of the C1C2 interval duration.

As Table 6 show, neither prediction was supported by the data. V duration in closed syllables was negatively correlated to C1C2 duration to the same extent as in open syllables, thus indicating that the hypothesized compensatory effect, though moderately present, was equally distributed over VCC sequences with different syllabic structures. Unfortunately we did not obtain any indication concerning the behavior of so-called «undecidable» clusters, the result of the test being probably unreliable due to the insufficient number of observations for this subgroup of data.

Concerning the second prediction, V duration was not correlated with C1 duration, both when we considered all clusters together ($N = 197$; $\rho = 0.17$, $p > .05$) and the three subgroups separately (Table 6). This result indicated therefore that, in sequences with heterosyllabic as well as tautosyllabic clusters, the duration of a stressed nucleus was completely unrelated to the duration of the individual C following it. On the contrary, it tended to be influenced by the duration of the whole C interval.

	Tautosyllabic clusters	Heterosyllabic clusters	«Undecidable» clusters
Pearson's ρ V-C1C2 correlation	-0.26**	-0.32**	n.s.
Pearson's ρ V-C1 correlation	n.s.	n.s.	n.s.
N	70	100	27

Table 6. Correlation coefficients (Pearson's ρ) between stressed V duration and the duration of the following consonantal interval (C1C2) or consonant (C1) (Corpus A).

4.2. Corpus B

As previously stated, Corpus B is a very large data sample, but it is also «noisier» in the sense that, on the one hand, annotation was semi-automatic (e.g., automatically segmented and then manually re-checked) and, on the other, the possibilities of selecting subsets of data with homogeneous characteristics are rather limited. We wanted to evaluate if the assumed compensatory forces within V-to-V intervals had an impact on large data samples, as the preliminary results on Mairano (2011) seemed to suggest, even though such data samples were not purposely constructed and included much structural variability (as specified below).

We therefore calculated the Pearson's correlation coefficient between the duration of all stressed as well as unstressed vowels, and the duration or number of segments of the following consonantal interval. It must be emphasized that «stressed» here means that the vowels carried lexical stress (according to the phonological rules of Italian stress assignment), but not necessarily an accent, as accents were not annotated in this corpus.We excluded from the computation vowels with duration < 20 ms on the assumption that in such cases some kind of error or segmentation imprecision might have occurred. Another important difference between Corpus B and Corpus A resided in the fact that, in Corpus B, consonantal intervals were not exclusively represented by a singleton consonant, a geminate, or two consonants. Instead, the composition of the consonantal intervals ranged from 0 (if a vowel was followed by another vowel, in cases of hiatus) to 4 segments (in rare cases and across word boundaries, such as *sto**p str**adale* 'road stop'). In this sense, Corpus B included much more structural heterogeneity than Corpus A.

The results of the correlation tests over the whole corpus are summarized in Table 7. In general, the coefficient values were extremely low, suggesting that almost no correlation could be found in the data. Suprisingly, unstressed Vs turned out to show a higher correlation coefficient than stressed Vs, in both cases of duration and number of Cs. In the case of stressed vowels, there was substantially no correlation with either the duration or the segmental length of the following consonantal interval.

	all vowels	stressed only	unstressed only
Pearson's ρ (duration of C interval)	-0.05**	+0.02**	-0.25**
Pearson's ρ (n. of Cs)	-0.13**	-0.10**	-0.28**
N	204900	63558	141343

Table 7. Correlation coefficients (Pearson's ρ) between V duration and the duration/number of the following Cs (Corpus B).

This results clearly reflected the problematic decision of basing the analysis on lexical stress, rather than on «true» accents. It seems plausible to hypothesize that, most frequently, unstressed Vs are also unaccented; in any case, for unstressed Vs, the fact of bearing or not bearing an accent would not imply significant changes in duration. On the contrary, the presence vs. absence of an accent for «lexically» stressed vowels is likely to make a noticeable difference for duration (as well as for other supra-segmental factors). As a consequence, analyzing Vs by lexical stress rather than by accent has more disruptive effects for the subgroup of stressed Vs than for the unstressed ones. This should then explain the fact that we found nearly no correlation for the latter category.

As a following step, we tried to reduce the number of potentially unaccented vowels from the stressed group by eliminating stressed Vs with duration < 70 ms. As shown in Table 8, the results improved slightly.

The analysis presented in the following section concerned only the selection of data in which stressed vowels had a duration >= 70 ms.

	all vowels	stressed only	unstressed only
Pearson's ρ (duration of C interval)	-0.06**	-0.06**	-0.25**
Pearson's ρ (n. of Cs)	-0.13**	-0.18**	-0.28**
N	194811	53465	141343

Table 8. Correlation coefficients (Pearson's ρ) between V duration and the duration/number of following consonants for data (stressed V >= 70ms) (Corpus B).

4.2.1. Analysis by speaker

Similarly to what has been done for Corpus A, we analyzed the production of each speaker of Corpus B separately. The results are shown in Table 9.

speaker		stressed only	unstressed only
F1	Pearson's ρ (duration of C interval)	+0.03**	-0.25**
	Pearson's ρ (n. of Cs)	-0.13**	-0.29**
	N	31978	87777
F2	Pearson's ρ (duration of C interval)	-0.18**	-0.31**
	Pearson's ρ (n. of Cs)	-0.30**	-0.33**
	N	11279	26879
M1	Pearson's ρ (duration of C interval)	-0.21**	-0.20**
	Pearson's ρ (n. of Cs)	-0.25**	0.22**
	N	10208	26687

Table 9. Correlation coefficients (Pearson's ρ) between V duration and the duration/number of following consonants split by speaker (Corpus B).

The values confirmed the presence of a weak but diffused negative correlation between the duration of the V and the duration of the subsequent C interval (ranging -0.20 to -0.31), as well as between the duration of the V and the number of consonants composing the following cluster (ranging from -0.22 to -0.33). However, one speaker (F1) departed from this trend in the case of the stressed Vs. The effects of inter-speaker variability that we observed for Corpus A were therefore confirmed for Corpus B as well. Recall that the speakers of Corpus B were all professionals and, even to a trained phonetician's ear, their regional origin was not detectable. Evidently there must be timing characteristics that are below the threshold of conscious control (and possibly of perception too). Further investigation of individual timing characteristics would be needed.

4.2.2. Analysis by number of consonants

As previously said, Corpus B included consonantal clusters composed of an unrestricted number of C, ranging from 0 (e.g. in the case of a hiatus) to 4. We restricted the by-speaker analysis to those intervals composed of 1 to 2 consonants. This sub-selection of data was therefore structurally very similar to Corpus A.

The results are shown in Table 10. Although the values of the Pearson's correlation coefficient did not change very much overall, what emerged straightforwardly in this sub-selection of data was the difference between stressed and unstressed Vs. Only the duration of the former decreased (though very moderately) when the duration or the number of segments composing the C interval increased. This result was more consistent with the general view according to which only stressed nuclei are influenced by the composition of V-to-V intervals; however, the timing effects acting at the level of unstressed vowels (consider in particular the difference between Table 10 and Table 9) remain completely obscure.

		stressed only	unstressed only
F1	Pearson's ρ (duration of C interval)	-0.05**	+0.05**
	Pearson's ρ (n. of Cs)	-0.14**	+0.07**
	N	31978	87777
F2	Pearson's ρ (duration of C interval)	-0.10**	+0.03**
	Pearson's ρ (n. of Cs)	-0.20**	+0.05**
	N	11279	26879
M1	Pearson's ρ (duration of C interval)	-0.10**	+0.07**
	Pearson's ρ (n. of Cs)	-0.20**	+0.09**
	N	10208	26687

Table 10. Correlation coefficients (Pearson's ρ) between V duration and the duration/number of following consonants split by speakers, C intervals = C or CC only (Corpus B).

5. Final remarks

In §2 we reviewed a few previous studies that showed the importance of accounting for V-to-V intervals in the explanation of the prosodic properties of Italian. We suggested that such evidence was mostly disregarded in successive research, and specifically we hypothesized that a syllable-independent timing effect exists in Italian, by which the duration of a stressed nucleus is inversely correlated to the duration of the consonantal interval following it. In §3 and §4 we presented the results of a production experiment aimed at testing this hypothesis. The results are still too partial to allow coherent arguments either supporting or denying the validity of the hypothesis. However, the trends discussed above suggested that such effect was present under certain conditions, although it could not alone account for all the variability attested in the data.

A first source of independent variation appeared to be the individual. Although all the subjects of Corpus A were asked to produce a relatively limited amount of highly controlled read speech, and those of Corpus B were professional speakers, there were substantial differences in the individual outputs as far as the relation between V and C duration in the target stimuli were concerned. Moreover, the two subjects of Corpus A who did not conform to the majority pattern were speakers of a southern variety of Italian, namely a southern Lazio and a Campanian variety (Sobrero 1988: 732). Their production showed a complete absence of correlation between stressed Vs and subsequent C intervals. This result therefore suggested the possibility that the first source of variation in our data was not (only) the speaker but (also) the linguistic variety. The issue of speech timing characteristics and diatopic differentiations is still almost unexplored in current research on Italian (see Turchi & Bertinetto 2000 for a similar concern), but a systematic investigation is likely to uncover interesting patterns of individual, local and supra-local convergence / divergence within the panorama of regional speech varieties («italiani regionali»; Sobrero 1988; Telmon 1990), that still are poorly investigated from a quantitative point of view.

The second source of variation is related to the quality of the stressed nuclei. Front mid vowels turned out to be consistently unaffected by compensatory shortening as a function of the subsequent C interval duration in Corpus A. Actually we cannot be entirely sure that this is a truly vowel-dependent effect; the possibility in fact exists that this is rather an idiosyncratic behavior of a certain group of target words (or sentences).

Differently from previous studies on V shortening in Italian, in which the experimental vowel was generally kept constant,[10] we tried to incorporate as much structural variation as possible concerning the quality of the stressed V. This however produced some variability in the durational patterns of the segments that we could not control for, due to the limited number of repetitions at our disposal. Future analyses of vowel onset intervals should therefore incorporate vowel quality variations in a more systematic way; this would permit to evaluate if different vowels are differently affected by timing phenomena and, in the positive case, to uncover the (phonetic) motivations for it.

Finally, although the data of Corpus A were normalized according to the average syllable duration in each sentence (and thus accounted for speech rate variations across speakers and sentences, at least to some extent), much residual variability proved impossible to be accounted for. This had to do with within-sentence speech rate variations (i.e., even within a short sentence there may be words uttered more slowly / emphatically than others), the distribution of accents (lexical stress does not automatically predict prosodic prominence, particularly in large data sample of «real» speech), lexico-syntactic idiosyncrasies (each word and word juncture being characterized by its own frequency of occurrence / plausibility)[11] and also segment-specific characteristics. Concerning the latter point, consider for example that intervocalic /r/ in Corpus A had an average duration of around 14 ms, while the /ps/ cluster was well above 250 ms. Since (stressed) vowels have, all things being equal, their own intrinsic durations as well as their own intrinsic limits of stretching and compressibility, such large variations in the consonantal intervals will obviously not affect vocalic variations proportionally.

In conclusion, the relationship between the vocalic and the consonantal slot within V-to-V intervals, though present in the data, was found to be weak to moderate, and unable to account for a large part of the durational variability of segments. Our data neither lent clear support to the alternative hypothesis of rhythmic differences exclusively associated to the opposition between open and closed syllables, though. In particular, in the case of open syllables followed by tautosyllabic onset clusters (Corpus A), the duration of the V in the first syllable was found to be inversely correlated to the duration of the onset of the subsequent syllable. This effect cannot be explained by referring to the phonological syllable as the primary timing unit in Italian. The fact that, in the case of heterosyllabic clusters, V duration was correlated (though moderately) with the duration of the entire C interval, rather than with the duration of the coda C, equally lent support to the notion of rhythmical syllable as the fundamental timing unit (Farnetani & Kori 1986).

Consistently with recent trends in rhythm studies, showing that linear effects of the phonological variables are not enough to explain the total variance in durational measures (Loukina et al. 2013), this paper has suggested that some low-level timing effects generated within voice onset intervals may account for part of this variability and are of a compensatory nature. The precise weight of compensatory factors in shaping the rhythm of (different varieties of) Italian, as well as the more general question of how

---

[10] Only /a/-stimuli were used by Farnetani & Kori (1986), Turchi & Bertinetto (2000) (and most works cited therein), Pickett et al. (1999) while in Marotta (1985) and (1995) the stressed vowel could be /a/, /o/ or /e/.
[11] Word frequency notoriously affects speech production, as reflected in the phonetic properties of the segments; see e.g. the existing evidence on vowel reduction (e.g. Van Coile 1987) or assimilation (e.g. Ernestus et al. 2006).

gradual compensatory effects can be integrated into categorical, i.e. phonological syllable-based constraints, remain open for future investigation.

6. Bibliography

ABERCROMBIE, David (1967). *Elements of General Phonetics.* Edinburgh: University Press.

ARVANITI, Amalia (2012). «The usefulness of metrics in the quantification of speech rhythm». *Journal of Phonetics* 40: 351-373.

BARBOSA, Plinio A. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes.

BARRY, William& ANDREEVA, Bistra& KOREMAN, Jacques (2009). «Do rhythm measures reflect perceived rhythm?». *Phonetica* 66: 78-94.

BERTINETTO, Pier Marco (1977). «"Syllabic Blood", ovvero l'italiano come lingua ad isocronismo sillabico». *Studi di Grammatica Italiana* 6: 69-96.

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

______ (1983). «Ancora sull'Italiano come Lingua ad Isocronia Sillabica». In: *Scritti linguistici in onore di Giovan Battista Pellegrini*. Pisa, Pacini: 1073-1082.

______ (2004). «On the undecidable syllabification of /sC/ clusters in Italian: Converging experimental evidence». *Italian Journal of Linguistics* 16: 349-372

______ (2011). «Sillaba». In: *Enciclopedia dell'Italiano*. Istituto dell'Enciclopedia Italiana Fondata da Giovanni Treccani S.p.A. http://www.treccani.it/enciclopedia/sillaba_(Enciclopedia_dell'Italiano)/

BERTINETTO, Pier Marco & BERTINI, Chiara (2008). «On modeling the rhythm of natural languages». In: Plinio A. Barbosa& Sandra Madureira & César Reis, eds.,*Speech Prosody 2008,Fourth Conference on Speech Prosody, May 6-9, Campinas, Brazil*.São Paulo, Capes: 427-430.

BERTINI, Chiara& BERTINETTO, Pier Marco& ZHI, Na (2011). «Chinese and Italian speech rhythm. Normalization and the CCI algorithm». In: *12thannualconferenceof theInternationalSpeechCommunicationAssociation2011(INTERSPEECH2011): Florence, Italy, 27-31 August 2011*. Red Hook, NY, Curran: 1853-1856.

BRADY, Michael C.& PORT, Robert F.& NAGAO, Kyoko (2006). «Effects of speaking style on the regularity of mora timing in Japanese». *Journal of the Acoustical Society of America* 120(5): 3208.

CELATA, Chiara & CALAMAI, Silvia (2011). «Timing in Italian VNC sequences at different speech rates». In: *12thannualconferenceof theInternationalSpeechCommunicationAssociation2011(INTERSPEECH2011): Florence, Italy, 27-31 August2011*. Red Hook, NY, Curran: 405-408.

CELATA, Chiara & KAEPPELI, Barbara (2003). «Affricazione e rafforzamento in italiano: alcuni dati sperimentali». *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore* 4.http://linguistica.sns.it/QLL/QLL03/CCelata.pdf

CHEN, Matthew (1970). «Vowel length variation as a function of the voicing of the consonantal environment». *Phonetica* 22: 129-159.

DELLWO, Volker& FOURCIN, Adrian& ABBERTON, Evelyn (2007). Rhythmical classification of languages based on voice parameters». In: Jürgen Trouvain & William Barry, eds., *Proceedings of the XVI International Congress of Phonetic Sciences: ICPhS XVI,* 6-10 August 2007*, Saarbrücken, Germany*.Saarbrücken: Univ. des Saarlandes:1129-1132.

D'IMPERIO, Mariapaola & ROSENTHALL, Sam (1999). «Phonetics and phonology of main stress in Italian». *Phonology* 16: 1-28.

ERNESTUS, Mirjam& LAHEY, M. & BAAYEN, Harald R. (2006). «Lexical frequency and voice assimilation». *Journal of the Acoustical Society of America* 120(2): 1040-1051.

FARNETANI, Edda & KORI, Shiro (1986). «Effects of Syllable and Word Structure on Segmental Durations in Spoken Italian». *Speech Communication* 5: 17-34.

FOWLER, Carol A. (1977). *Timing Control in Speech Production*, Indiana University Linguistics Club.

HAJEK, John & STEVENS, Mary& WEBSTER, Georgia (2007). «Vowel duration, compression and lengthening in stressed syllables in Italian». In: Jürgen Trouvain & William Barry, eds., *Proceedings of the XVI International Congress of Phonetic Sciences: ICPhS XVI,* 6-10 August 2007*, Saarbrücken, Germany*. Saarbrücken: Univ. des Saarlandes: 1057-1060.

HERMES, Anne (2013). *Articulatorycoordination and syllable structure in Italian*. Frankfurtam Main: Peter Lang.

HIRATA, Yukari & FORBES, Connor (2007). «Vowel-onset interval as a timing unit for singleton/geminate stop distinction in Japanese». In: Jürgen Trouvain & William Barry, eds., *Proceedings of the XVI International Congress of Phonetic Sciences: ICPhS XVI,* 6-10 August 2007*, Saarbrücken, Germany*. Saarbrücken: Univ. des Saarlandes: 349-354.

LINDBLOM, Bjorn & Rapp, Karin (1973). «Some temporal regularities of spoken Swedish». *Papers in Linguistics from the University of Stockholm* 21: 1-59.

LOUKINA, Anastassia & Rosner, Burton& Kochanski, Greg& Keane, Elinor& Shih, Chilin (2013). «What determines duration-based rhythm measures: text or speaker?»*Laboratory Phonology* 4: 339-382.

MAIRANO, Paolo (2011). *Rhythm typology: acoustic and perceptive studies.*PhD Thesis, University of Turin.

MAROTTA, Giovanna (1984). *Aspetti della struttura ritmico-temporale in italiano. Studi sulla durata vocalica*. Pisa: ETS.

______ (1985). *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna: Zanichelli.

______ (1995). «La sibilante preconsonantica in italiano: questioni teoriche ed analisi sperimentale». In: Roberto Ajello & Saverio Sani, eds., *Scritti linguistici e filologici in onore di Tristano Bolelli.* Pisa, Pacini: 393-438.

MOOSMÜLLER, Sylvia & RINGEN, Catherine (2004). «Voice and aspiration in Austrian German Plosives». *Folia Linguistica* 38: 43-62.

MOOSMÜLLER, Sylvia & BRANDSTÄTTER, Julia (2014, in press). «Phonotactic information in the temporal organisation of Standard Austrian German and the Viennese Dialect». *Language Sciences*.

PICKETT, Emily& BLUMSTEIN, Sheila& BURTON, Martha W. (1999). «Effects of speaking rate on the singleton/geminate distinction in Italian». *Phonetica* 56: 135-157.

PIKE, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

PORT, Robert F. (1979). «The influence of tempo on stop closure duration as a cue for voicing and place». *Journal of Phonetics* 7: 45-56.

SOBRERO, Alberto A. (1988). «Italienisch: Regionale Varianten / Italiano regionale». In: Günter Holtus, Michael Metzeltin & Christian Schmitt, eds., *Lexikon der romanistischen Linguistik (LRL)*, Vol. IV. Tübingen, Niemayer: 732-748.

TELMON, Tullio (1990). *Guida allo studio degli italiani regionali*. Alessandria: Edizioni dell'Orso.

TILSEN Sam & ARVANITI, Amalia (2013). «Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages». *Journal of the Acoustical Society ofAmerica* 134(1): 628-39.

TURCHI, Laura & BERTINETTO, Pier Marco (2000). «La durata vocalica di fronte ai nessi /sC/: un'indagine su soggetti pisani». *Studi Italiani di Linguistica Teorica e Applicata* 29: 389-422.

VAN COILE, Bert (1987). «A model of phoneme durations based on the analysis of a read Dutch text». In: *Proceedings of the European Conference on Speech Technology, Edinburgh, Scotland, UK, September 1987*. Edinburgh:vol. II, 233-236.

VAN SANTEN, Jan P. H.& COLEMAN, John S.& RANDOLPH, Mark A. (1992). «Effects of postvocalic voicing on the time course of vowels and diphthongs». *Journal of the Acoustical Society of America* 92 (4): 2444.

VAN SANTEN, Jan P.H. & SHIH, Chilin (2000). «Suprasegmental and Segmental Timing Models in Mandarin Chinese and American English». *Journal of the Acoustical Society of America* 107(2): 1012-1026.

VAYRA, Mario& AVESANI, Cinzia& FOWLER, Carol (1999). «On the phonetic basis of vowel-consonant coordination in Italian: a study of stress and 'compensatory shortening'». In: John J. Ohala, ed.,*Proceedings of the XIV International Congress of Phonetic Sciences, ICPhS 99, San Francisco, 1-7 August 1999*.Berkeley, Calif., University of California: 495-498.