Timing in Italian VNC sequences at different speech rates

Chiara Celata [1], *Silvia Calamai* [2]

[1] Scuola Normale Superiore, Pisa, Italy
[2] Università di Siena - Arezzo, Italy
c.celata@sns.it, calamai@unisi.it

Abstract

This study addresses the question of temporal cohesion in Italian word-medial VNC (vowel-nasal-obstruent) sequences varying in the laryngeal status of the post-nasal consonant, for two classes of obstruents distinct in terms of place and at three different speech rates. The temporal relations among the obstruent and the two preceding sonorant segments are examined, and variations in speaking tempo are shown to affect the timing pattern of different speech units in different ways. These results support a view of speech timing control in which temporal effects over constituents spanning syllable boundaries are to be combined with the effects observed over traditional syllable-sized units.

Index Terms: speech timing, speaking rate, nasal obstruent clusters, cluster internal balance.

1. Introduction

Speech timing control patterns are of particular interest to the study of language because they are at the crossroads of cognition and the mechanics of speech production. Sequences of vowels and consonants are temporally organized in a complex way which reflects both universals of speech articulation (grounded in the physiology of vocal tract constraints), and language-specific phonetic and phonological factors. Much of the research on these topics has been conducted on English and other Germanic languages.

This study focuses on intrinsic timing in Italian sequences of stressed vowels followed by homorganic nasal+stop clusters. It has been known for at least seventy years that the duration of stressed vowels is influenced by the nature of the surrounding consonants. In particular, a vowel tends to be longer when followed by a voiced consonant than when followed by a voiceless consonant, whereas the nature of the preceding consonant does not appear to have an influence to the same extent (see [9] and references therein, which date the observation of similar effects in English back to the 1940s; see also [21,12,13,18]; but see [4] for more cautious observations about the strength of the effect across phonotactic contexts and speech styles). Though no strong evidence exists on whether the effect of consonant voicing on the duration of the preceding vowel is exerted by syllable-initial and syllable-final consonants to the same extent, it is generally said that the effect is stronger for consonants in coda of the same syllable, that is, when the vowel and the consonant do not span a syllable boundary (e.g., [12] explicitly states that the effect of the postvocalic context is greatest "at phrase and clause boundaries" in English). Moreover, this effect of (mostly word-final) consonant voicing on vowel duration is associated with a general tendency for word-final voiceless obstruents to be longer than their voiced cognates (e.g., [16,22,23]; see also [17] for initial stops). Taken together, these data suggest that a compensatory relationship exists between vowel duration and closure duration, providing a cue to voicing [23,18]; moreover, compare the general tendency observed in many languages to shorten vowels before phonetically or phonologically long consonants (e.g., [5] for Swedish, [15] for Estonian, [8,1] for Italian etc). Finally, variation in pre-consonantal vowel duration has also been shown to function in perception as a cue for discriminating voiceless vs voiced final consonants (e.g., [25,24]). Besides English, French also seems to be a language in which a strong lengthening-before-voicing effect appears [20].

Focusing on the articulatory characteristics of stops differing in voicing, [10] provided some interesting data (though not completely conclusive, as the authors themselves declared) suggesting that voiceless and voiced stops may be distinguished by different patterns of coarticulation not only in post-vocalic position, but also in the context of a preceding or following nasal (i.e., in both contexts VCNV and VNCV). To our knowledge, the first two studies to investigate the effects of stop voicing on the preceding nasal are [3,27], both on English. According to their data, nasals before /d/ (as in *tending*) are longer than nasals before /t/ (as in *tenting*), thus apparently mirroring the behavior of vowels in the same contexts. This is not a completely new finding: some mention in the literature had already been made of a possible relationship between nasal duration and laryngeal specification of the post-nasal in NC clusters (e.g., [19,6]) and, recently, similar timing effects in clusters of nasal + alveolar stop (and to a lesser extent, nasal + velar stop as well) have been proved to be the source of a variety of nasal place assimilation processes in some Italo-Romance varieties [2].

[3,27] provided data on the duration of the preceding vowel as well. These two studies, both on English, differ, however, with respect to the results obtained. [3] showed that before a voiced obstruent the nasal and the vowel were both lengthened compared to the same categories before a voiceless obstruent. This effect was reinforced by the results of [4], also on English. By contrast, in [27] vowel duration was not affected by the voicing contrast in the post-nasal consonants, and the authors concluded that the difference in nasal duration appears to be a much more reliable cue to the voicing contrast between /t/ and /d/ than the durational difference in the preceding vowel (see also [13]). Interestingly, a radically different picture emerged from [8] on Italian disyllables with medial NC clusters. In this language, only in the case of nasal + voiced consonants were stressed vowels shorter than final vowels, while much variation was found for vowels preceding clusters of voiceless consonants. On the basis of this empirical picture, though unclear, [26] concluded that, within Italo-Romance, the shorter the nasal, the longer the vowel.

Such inconsistency in cross-linguistic experimental evidence for pre-nasal vowels is intriguing not only *per se*, but also if one considers the motivations that are adduced by the authors to explain either the absence or the presence of a

timing effect within the VNC constituent. On the one hand, both those who must explain the observed consistency between nasal and pre-nasal vowel lengthening before voiced obstruents and those who deny the existence of vowel lengthening across nasals motivate this fact by assuming that the stop voicing contrast bears (or does not bear) an effect on the preceding vowel, and on all possible intervening segments [27,13]. On the other hand, to explain data such as those discussed for Italian, where an inverse temporal relationship is said to hold between the nasal and the pre-nasal vowel [26], the tendency is to relate the temporal behavior of pre-nasal vowels to a difference in syllabic constituency. In particular, it is assumed that pre-consonantal nasals effectively close the syllable only when they precede a voiced stop, while they would have less of an effect when they precede a voiceless stop, because of their contextual shortening. Since Italian is a language in which vowels in open syllables tend to be longer than vowels in closed syllables (e.g., [8,7]; but, *contra*, see [14]), a different behavior of nasals with respect to syllable structure would automatically generate a difference in the duration of the vowel [26]. In this view, then, the element which directly influence the temporal behavior of the vowel is the nasal and not the post-nasal consonant.

In conclusion, it remains an open question whether a vowel in a VNC sequence is affected by timing constraints springing from the following segments, and, if such constraints do exist, whether it is the nasal or the post-nasal that generates them. In addition, it is not clear whether only word-final tauto-syllabic consonants (such as *bank* vs *bang* in English) have an effect on the temporal relation with the preceding segment, or this effect can be generalized to word-medial and heterosyllabic positions as well (such as *cantano* vs *mandano* in Italian).

This study addresses the question of VNC intrinsic timing relative to the laryngeal status of the post-nasal consonant for Italian word-medial positions at three different speech rates and for two types of clusters differing in the place of articulation of the post-nasal consonant (alveolar vs velar). The temporal relations between a voiceless or voiced post-nasal consonant and the preceding nasal, and between a voiceless or voiced post-nasal consonant and the preceding stressed vowel, are examined. The timing effects were evaluated against three different speech rate conditions, in the hypothesis that variations in speaking tempo may affect both absolute durations and relative timing patterns of speech units in a non-linear way. Indeed, previous investigations ([2]) demonstrated that different voicing characteristics in nasal homorganic clusters influenced the cluster internal balance with respect to the relationship between the nasal and the post-nasal consonant in the slow speech of the speakers to a greater extent than in the normally or fast uttered speech, and this discrepancy had implications for the reconstruction of different patterns of nasal place assimilation within the cluster. We therefore included a stylistic variable in the experimental design by analyzing three varieties of read speech; we hypothesized that certain divergences reported in the literature may partly be accounted for in terms of different degrees of permeability to timing restructuration for different stylistic conditions.

2. Experimental design

Sixteen meaningful Italian words containing a nasal-to-alveolar and a nasal-to-velar cluster were embedded in short isosyllabic frame sentences. The postnasal stop could either be voiceless (/nt/, /nk/) or voiced (/nd/, /ng/). Target words were trisyllables (e.g., *cantano*, *mandano*, *mancano*, *vangano*) selected in order to preserve an invariable vowel context (/anCa/); lexical stress was always on the first syllable of the word. Four native Italian speakers with no reported speech, language or hearing pathology, aged 30-35, speaking a Tuscan variety of Italian were recorded separately in an anechoic chamber. Subjects were naïve to the purpose and design of the study. The experimental sentences were randomized and the participants were asked to produce 30 repetitions of each sentence, 10 at a normal rate of speaking, 10 at a slow rate, and 10 at a fast rate. For normal speech elicitation, subjects were asked to produce 'normally' uttered speech with the target sentences being preceded by contextual questions. For slow speech elicitation, they were asked to produce clear uttered speech as if they were speaking to non-native listeners. Finally, for fast speech elicitation, they were asked to produce fast uttered speech without clipping syllables or single sounds.

The duration of both whole sentences and individual phonemes within the /anCa/ sequence was measured according to general segmentation criteria. Then the duration ratios between the post-nasal consonant (closure duration only) and the preceding nasal and between the post-nasal consonant and the pre-nasal stressed vowel was calculated. Both raw durations and duration ratios were taken as dependent variables and the experimental factors were Rate (normal vs fast vs slow), Place of articulation of the post-nasal consonant (alveolar vs velar) and Voicing of the post-nasal consonant (voiced vs voiceless). Analyses of variance and linear correlation coefficients were calculated (see below).

3. Analysis and discussion

The three speech rates were significantly different for sentence duration ($p < .0001$), /anCa/ duration ($p < .0001$) and /nC/ cluster duration ($p < .0001$), thus confirming that the three elicited stylistic conditions were temporally distinguished even when we considered speech units of different levels.

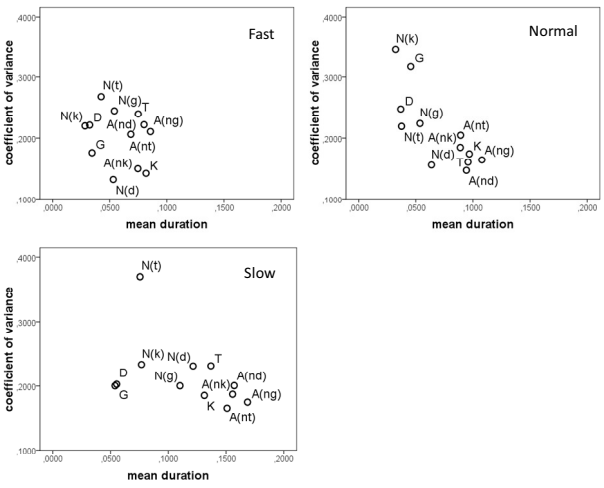

Figure 1: *Italian segmental durations from multiple speakers' data. Mean durations for each phone (sec) are plotted against the coefficient of variance.*

Variation was observed not only among phone classes, but also according to speaking styles. The cumulative action of contextual and stylistic variation gave rise, as expected, to a

complex pattern of temporal relations among sounds in sequence. Figure 1 shows the durational distributions from the production of the four native Italian speakers of this study. The figure plots mean segmental duration against the coefficient of variance (i.e., the standard deviation of each phone as a proportion of its mean) for each phone class and for the three speech styles (capital letters are used to identify the phones and small letters are used to identify the contextual segments). The figure shows that the relative variance was not uniform but tended to increase in the normal and the slow speech styles with respect to the fast speech, independently from the phone mean duration, and is generally higher for the nasals than for vowels (and stops). Voiced stops exhibited more variance than their voiceless counterparts, especially in the normal speech, while the reverse pattern was observable for the pre-consonantal nasals, with pre-voiceless nasals often being more variant than pre-voiced nasals. Vowels were generally clustered together for both mean duration and variance within each style, and they unsurprisingly appeared to be lengthened in a consistent way as speakers moved from fast to normal and to slow speech.

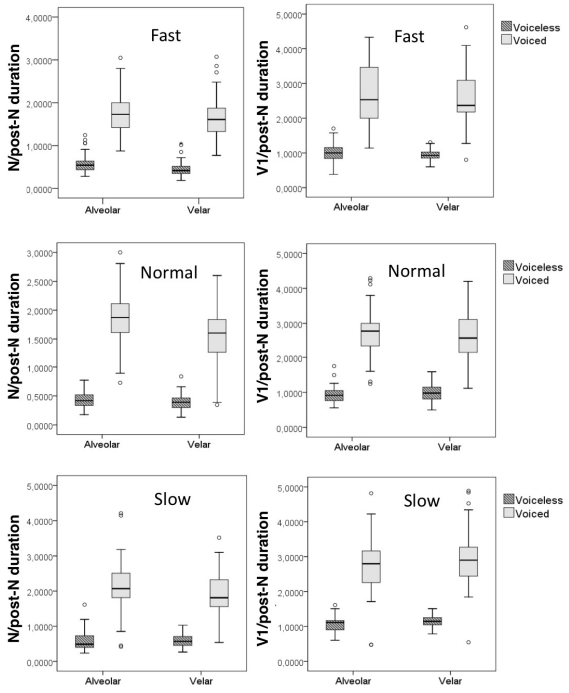

Figure 2: *Segmental duration ratios split for voicing and place of articulation of the post-nasal consonant, for the three speaking styles separately. Graphs on the left: nasal to post-nasal duration ratio; graphs on the right: V1 to post-nasal duration ratio.*

In order to evaluate the strength of the timing relations within the VNC unit, the interaction between post-nasal voicing and the relative duration of the preceding segments was analyzed, for the three speech rates and the two places of articulation. An analysis of variance with Voicing as the independent factor was carried out on duration ratios between the nasal and the post-nasal consonant and between the preceding stressed vowel (henceforth, V1) and the post-nasal consonant. For both variables, the results were consistent across speech rates and contexts (see Figure 2). The duration ratio between the nasal and the post-nasal was significantly longer for voiced clusters (/ng, nd/) than for voiceless ones (/nk, nt/) ($p = .000$ in all cases), thus indicating that in the voiced condition the nasal occupied a larger portion of the cluster (with respect to the portion occupied by the post-nasal), than in the voiceless condition (Figure 2, left). This pattern was consistent with previous results obtained for Italian ([2]) and other languages (see references above). Importantly, the balancing effects held true to the same extent for the three speech rate conditions, and without significant distinctions between the velar and the alveolar context. A similar effect was found for the preceding vowel (Figure 2b). The duration ratio between V1 and the post-nasal consonant was significantly greater for voiced clusters than for voiceless ones, indicating that the stressed vowel is much longer than the post-nasal stop when the latter is voiced, with respect to the case in which the post-nasal is voiceless. This result exactly parallels the timing pattern attested for the relationship between the nasal and the post-nasal, closely resembling the temporal behavior of pre-obstruent vowels and nasals found by [3,4] for English. For the same reason, this result is not consistent with the observation that word-final obstruent voicing does not exert any effect on pre-nasal vowel duration ([27]). Moreover, the presumed tendency of vowels to shorten before long nasals in clusters of voiced obstruents, and to be shorter before voiceless obstruents ([26]), appears to be completely unsupported in our data.

Given that, according to data in Figure 1 above, voiceless obstruents (/t k/) were overall longer than the corresponding voiced ones (/d g/) in normal speech, and this difference was maintained at fast and slow speech rates as well, we wanted to investigate further whether and to what extent the duration of the preceding segments varied as a function of obstruent voicing. Figure 1 is rather uninformative on this point, since for some of the preceding vowels and nasals the variance in duration increased considerably. Therefore, a Pearson's correlation coefficients between the nasal and the post-nasal duration values and between the V1 and the post-nasal duration values was calculated for the three styles separately (Figure 3). As far as the nasal was concerned, the correlation turned out to be negative and strongly significant, thus confirming that the longer the stop, the shorter the nasal (Figure 3, graphs on the left). The effect was stronger for the normal ($r = -.737$, $p = .000$) and fast speech ($r = -.629$, $p = .000$) and slightly weaker for the slow speech ($r = -.430$, $p = .000$). On the other hand, the correlation between vowel duration and post-nasal duration was non-significant for the slow speech ($r = .030$, $p > .05$) and negative and strongly significant for the normal ($r = -.242$, $p < .005$) and fast speech ($r = -.237$, $p < .005$) (Figure 3, graphs on the right), though r coefficients were consistently lower than those calculated for the nasal-obstruent correlation. In other words, while the nasal changed its duration as a consequence of a balancing effect with respect to the obstruent duration, the preceding vowel did not change to the same extent. Moreover, it is worth noticing that the balancing effect was weaker for the slow speech, if compared to the normal and fast speech.

4. Conclusions

This study has demonstrated that, similarly to the timing effects found for English in word-final VNC sequences, in Italian word-medial sequences the laryngeal specification of the post-nasal consonant determines the relative duration of the two preceding segments in the sequences. The balancing effort with respect to the durational properties of the post-

nasal appears to be carried out mostly by the adjacent nasal, and to a lesser extent by the preceding vowel.

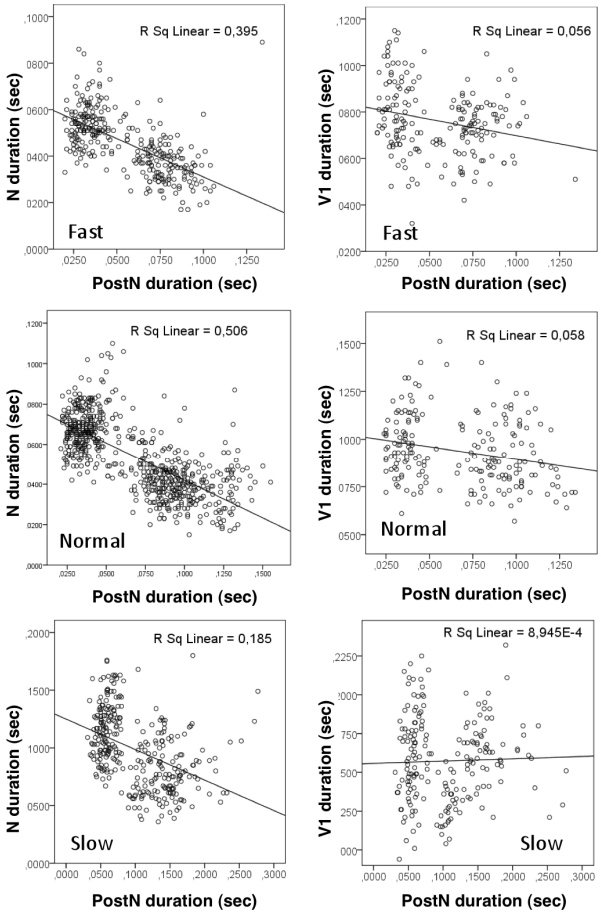

Figure 3: *Voiced and voiceless post-nasal consonant durations plotted against nasal durations (graphs on the left) and V1 durations (graphs on the right) for the three speaking styles.*

The timing effect was consistently represented across places of articulation and speaking styles, but the magnitude of the durational balance was greater for the normal and the fast speech styles with respect to the slow speech style. It can therefore be provisionally concluded that, if one takes the normal speech style as the baseline condition for the observation of timing effects among segments, slowing down the speaking tempo apparently perturbs the temporal cohesion of the speech units more than any departure from the baseline condition in terms of fastening. This asymmetric pattern strongly suggests, if one needs to stress further the importance of stylistic variation for any account of speech timing phenomena, that speaking rate variations must be included as a potentially relevant factor shaping the patterns of temporal control exhibited by the speakers.

The fact that word-medial hetero-syllabic sequences were found to show comparable effects than word-final ones has implications on the nature of the constituents to which the timing constraints turn out to apply. The present data show in fact that the voicing effect is strong enough to overcome the presence of a syllable boundaries within the VNC sequence. Syllable-timing hypotheses (such as in [11]), based on the observation that segmental durations can be well predicted by fitting them into a syllabe-level timing framework, should be integrated also in terms of timing effects spanning a syllable boundary.

References

[1] Bertinetto, P.M., Strutture prosodiche dell'italiano, Firenze, Accademia della Crusca, 1981.

[2] Calamai, S. and Celata, C., "Towards a sociophonetic explanation of progressive and regressive assimilation in Italian NC clusters", in *Proceedings of Sociophonetics, at the crossroads of speech variation, processing and communication*, Scuola Normale Superiore, Pisa, Italy, 2011 (in press).

[3] Chen, M., "Vowel length variation as a function of the voicing of the consonant environment", Phonetica 22:129-159, 1970.

[4] Crystal, T.H. and House A.S., "Segmental durations in connected-speech signals: Current results", JASA, 83:1553-1573, 1988.

[5] Elert, C.C., Phonologic Studies of Quantity in Swedish, Uppsala: Almqvist & Wiksell, 1964.

[6] Espinosa, A.M., Estudios sobre el español de Nuevo Méjico, Buenos Aires, 1930.

[7] Farnetani, E. and Kori S., "Lexical Stress in Spoken Sentences. A Study on Duration and Vowel Formant Pattern", Quaderni del Centro di studi per le ricerche di Fonetica, 1:104-133, 1982.

[8] Fava, E. and Magno Caldognetto, E., "Studio sperimentale delle caratteristiche elettroacustiche delle vocali toniche ed atone in bisillabi italiani", in R. Simone, U. Vignuzzi, A. Ruggiero [Eds], *Studi di fonetica e fonologia*, 35-79, Roma, Bulzoni, 1976.

[9] House, A.S. and Fairbanks, G., "The Influence of consonant Environment Upon the Secondary Acoustical Characteristics of Vowels", JASA, 25:105-113, 1953.

[10] Kent, R.D. and Moll, L., "Vocal-tract characteristics of the stop cognates", JASA, 46:1549-1555, 1969.

[11] Campbell, W.N. and Isard, S.D. "Segment durations in a syllable frame", J. Phon. 19: 37-47, 1991.

[12] Klatt, D.H., "Synthesis by rule of segmental durations in English sentences. In B. Lindblom & S. Öhman [Eds], *Frontiers of Speech Communication Research*, 287-299, New York, Academic Press, 1979.

[13] Klatt, D.H., "Review of text-to-speech conversion for English", JASA, 82:737-793, 1987.

[14] Landi, R., and Savy, R., "Durata vocalica, struttura sillabica e velocità di eloquio nel parlato connesso", in A. Peretti and P. Simonetti (Eds), *Atti del XXIV Convegno Nazionale dell'Associazione Italiana di Acustica*, 65-70, Padova, 1996.

[15] Lehiste, I., Consonant quantity and phonological units in Estonian, Bloomington, Indiana University, 1966.

[16] Lisker, L., "Minimal cues for separating /wrly/ in intervocalic position", Word, 13:256-267, 1957.

[17] Lisker, L. and Abramson, A. S., "Some effects of context on voice onset time in English stops", Lang.Speech, 10:1-28, 1967.

[18] Luce, P.A. and Charles-Luce, J., "Contextual effects on vowel duration, closure duration, and the C/V ratio in speech production", JASA, 78:1949-1957, 1985.

[19] Mayer E., Englische Lautdauer, Uppsala, Lundeqvist, 1903.

[20] O'Shaughnessy, D., "A study of French vowel and consonant durations", J. Phonet. 9:385-406, 1981.

[21] Peterson, G.E. and Lehiste, I. "Duration of syllable nuclei in English", JASA 32:693-703, 1960.

[22] Port, R. F., "The influence of tempo on stop closure duration as a cue for voicing and place", J. Phonet., 7:45-56, 1979.

[23] Port, R.F., "Linguistic timing factors in combination", JASA, 69:262-274, 1981.

[24] Port, R.F., and Dalby, J., "C/V ratio as a cue for voicing in English", Percept. Psychophys 3.2:141-152,1982.

[25] Raphael L.J., "Durations and contexts as cues to word-final cognate opposition in English", Phonetica 38:126-147, 1981.

[26] Tuttle, E.F., "Nasalization in Northern Italy: syllabic constraints and strength scales as developmental parameters", Italian Journal of Linguistics, 3:23-92, 1991.

[27] Zue, V.W. and Laferrière, M., "Acoustic study of medial /t,d/ in American English", JASA, 66:1041-1050, 1979.