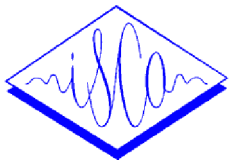

Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

On the phonetics of rhythm of Italian: patterns of duration in pre-planned and spontaneous speech

Rosa Giordano

Department of Linguistics and Literature
University of Salerno, Italy
rogiordano@unisa.it

Abstract

This paper is concerned with the phonetic correlates of main stress in Italian and presents the analysis of the duration of syllabic *nuclei* in connected speech and in different speaking styles. Spontaneous dialogues are analysed basing on earlier works on pre-planned speech which investigated both [+stress] / [-stress] and stress/accent contrasts in Italian.

The aim of this study is to examine the effects of two prosodic factors on the duration pattern of the intonation groups: 1) phonological prominences pertaining to different prosodic tiers (rhythmical and intonational), and 2) their position in the intonation group.

Results show that the temporal pattern of the intonation group would actually depend on different factors on both prosodic layers: stress, pertaining to rhythm, and post-lexical factors, pertaining to intonation and to prosodic phrasing.

1. Introduction

Italian, as other Romance languages, presents main stress, which in current research is intended to be the phonological characterization of a syllable as prominent with respect to the other syllables in the word [1][2][3]; the rhythmical positions marked by main stress are accepted to be the placeholders of pitch accents [2][4][5][6]. From a phonological viewpoint, stress and accent are similar, as they are both exponents of the culminative phonological property, although they pertain to different domains: stress is related to the lexical domain and accent is related to the post-lexical, or sentence, domain.

Several works tested the phonetic realization and the perception of lexical stress in Italian on laboratory speech and read speech, in controlled contexts, in some cases also evaluating the effects of intonational accents and boundaries (see, among others, [1][7][8][9][10][11][12][13]). Results show that stressed syllables entail longer mean durations than unstressed ones, usually present higher values of intensity and can present F0 variations if sentence accents are associated with them. According to such works, duration would be the main acoustic cue of stress in Italian, while frequency would be related to accent; studies on the articulatory properties of prominence in Italian also support this conclusion [14].

In Italian, duration is involved in phonological contrasts as for the segmental level [2]. In fact, length acts as distinctive feature for consonantal contrasts: *fato* [`fa:to] (destiny, fate) and *fatto* [`fatto] (fact) are different lexical items; vowels, and then syllabic *nuclei,* do not share this phonological property with consonants. Vowel length, instead, is automatically related to a) main stress, b) main stress position in the word and c) syllabic structure [1][2]. Connected spontaneous speech sometimes provides different evidences; recent works suggest that vowel length could be not actually sensitive to the syllabic tail, as in spontaneous speech productions vowel duration in open and closed syllables does not consistently differ [15][16][17][18].

Studies on the temporal variation of the syllabic components under different prosodic conditions show that lengthening mostly affects *nuclei* [7][19] (see [20] for syllabic theories).

This analysis is focused on the duration of syllabic *nuclei* in spontaneous speech, in order to test if stressed syllables phonetically differ from unstressed neighboring syllables and if their phonetic realization changes when sentence-level prominences are associated with them and when they occur at the edge of a prosodic group. Research on the interplay between temporal and melodic features [21-23] is also relevant to two main issues in the fields of prosodic phonology and of articulatory phonology, usually supported by findings coming from laboratory speech: 1) the relation between rhythm and intonation, 2) prosodic phrasing (e.g. [4:50 ff.][24][25]). From studies on the acoustic cues of stress and accent in Italian connected pre-planned speech [26][27], on which this study bases, some preliminary results emerged:

- duration seems to be a reliable acoustic parameter related to stress: in the speech stream, duration values of each stressed syllable are frequently higher than the values of the preceding unstressed syllable; intensity also follows the same alternation, probably supplying duration when temporal values do not increase;
- duration variability [15][16][17][18][26][28] does not necessarily contradict such evidence: in fact, though duration values of both stressed and unstressed syllables are directly correlated to speech rate, thus presenting wide absolute ranges of variation and of overlap, they systematically follow the dynamics cited above [26];
- accent would not systematically affect vowel duration;
- final position in the prosodic group entails additional lengthening of stressed or accented syllables; depending on the boundary tones they are associated with, also unstressed syllables can lengthen [27] (see also [29]).

2. Methods

2.1. Corpus

Two sets of connected speech representing different speaking styles have been compared: 1) News of the main Italian TV channels, 2) task-oriented dialogues selected from the national *corpus* CLIPS [30]; the former is a sample of pre-planned speech, the latter of phonetically spontaneous speech.

Set 1 consists of 4 streams of speech, each of them about 45” long, performed by 4 journalists, who are also professional native speakers. Set 2 consists of 2 elicited *spot the differences* dialogues, named DGtdA04O and DGtdB04R;

the 4 speakers are university students. The same material has been used in [26] and [27].

2.2. Prosodic analysis

Speech streams have been divided into prosodic groups [4:72 ff.][5][6], basing on instrumental and acoustic analysis, phonetic *criteria* and uditive coherence (final lengthening, prosodic cohesion and general trends of F0 and energy). Intonational patterns have been analysed in order to identify relevant pitch movements (in AM terms: accents or boundary tones), also taking into account descriptive works and different theoretical statements on the intonation of Italian (see, e.g., [1][2]). The [+accent] / [-accent] opposition and prosodic phrasing are then mainly empirically determined.
The [+stress] / [-stress] contrast is phonologically determined. For each prosodic group, strong and weak syllables have been marked on the rhythmical grid, following the phonology of rhythm of Italian [1][2]. So that the terms *stressed* and *unstressed* correspond here to rhythmical *strong* position and rhythmical *weak* position. The analysis concerns the vowel, which phonologically corresponds with the syllabic *nucleus*: syllables presenting diphthongs and cases of problematic segmentation are excluded.
In Italian the position of the main stress is not fixed in the word; considering factors affecting vowel length in Italian (see in particular item b) reported in §1), the subset of stressed syllables analysed in this work does not include stressed syllables occurring in the ultimate position into the word.
Final position is intended to be the rightmost edge of both major or minor prosodic groups and starts from the last strong syllable; in consequence of what stated before, then:

- final stressed syllables are all penultimate or antepenultimate syllables;
- final unstressed syllables are all post-tonic syllables.

Considering the evidence about the acoustic cues of stress and accent in Italian, reported above, in this work the effects of the syllabic structure on the length of the *nucleus* are not analysed in detail. Data about this topic are available in [26].

Words ending with syllables whose lengthening is due to emotional facts or to hesitation have been excluded, as well as words interrupted for resetting or for turn-taking by the other speaker (for details about fluency, see [28]).
Table 1 reports the number of syllables examined.

Table 1: *Number of syllables for each prosodic condition.*

		Prominence level			Total
		Unstressed	Stressed	Accented	
Position	Non final	2820	891	636	4347
	Final	625	42	569	1236
Total		3445	933	1205	5583

2.3. Acoustic measures

The sound files were analysed using Praat and Wavesurfer.

The prosodic groups were analysed in detail and manually labelled, using the INTSINT system of annotation and then getting the f0 value (in Hz) of each labeled target point [5]. Duration (in milliseconds) of head, *nucleus* and tail of each syllable was measured, using waveforms and spectrogram to guide segmentation and labeling (see [31] for notes about this topic). F0 and intensity (in dB) values were estimated over the stable part of each vowel.

3. Results and discussion

3.1. Prominence level

Mean durations of syllabic *nuclei* are shown in the graphs reported below. Data relative to *accented*, *stressed* and *unstressed* syllables, both in *final* and in *non final* position, are respectively plotted in figure 1, 2 and 3. Speakers are identified by a progressive number and with *D*, for dialogue speakers, and *tv*, for television journalists.

The first evidence concerns the range of variation of the values for each speaker and for each class considered.

As said above, previous analyses [25] proved that the phenomenon could be due to other prosodic factors, in particular to speech rate changes, and that a precise temporal dynamics corresponding to the rhythmical pattern actually occurs in the speech flow (see also the temporal model for other languages [32]): the majority of stressed and of accented (i. e. rhythmically *strong*) syllables are longer than the preceding unstressed (i.e. rhythmically *weak*) syllable.

However, this trend is not an absolute rule: in a consistent number of occurrences, strong syllables are not actually longer than the preceding syllable. In my opinion, this phenomenon can be explained considering the prosodic context and some production strategies: that is both structural and performance factors. In fact, all of these syllables are placed in so-called zones of *transition* [see, e.g., 4:45] between intonational events and are not directly associated with intonational events. Moreover, in unplanned and spontaneous speech speakers tend to hypoarticulation [33], which in turn implies phonetic reduction, acting as an important factor of variation. Phonetic segmental variability in hypospeech is a well-known fact and its correlation with the prosodic patterns has been clearly demonstrated for Italian [34][35]. The phenomenon highlighted here would now suggest the possibility that phonetic reduction, or undershoot, could also affect the prosodic production, besides the segmental one, and would well fit in with the results of some articulatory analyses (see [14] and references by the authors).

Vowel duration is someway related to the prominence level, as unstressed syllables (figure 3) always present lower values than accented syllables (figure 1) for all the speakers.

Anyway, the two speaking styles consistently differ.

As for TV speakers, the difference in the duration between rhythmical weak syllables and rhythmical strong syllables is clear, but unexpectedly stressed and accented vowels slightly differ in their values (compare figures 1 and 2). On the contrary, in dialogic speech the differences between stressed and unstressed syllables lessen (see fig. 2 and fig. 3), and in some case seem to be null, while accented syllables are consistently longer than stressed syllables.

Thus the main difference between the two styles regards the *accented* syllables set compared with the *stressed* syllables set. As for the *unstressed* set, duration values are strongly similar in unmarked conditions, i.e. in internal positions of the prosodic group.

Some of these characteristics can be explained as related to speaking styles: TV journalists generally tend to hyperarticulation and frequently use a particular degree of emphasis which can also be signalled by non-pitch features in several languages [36] and probably also in Italian.

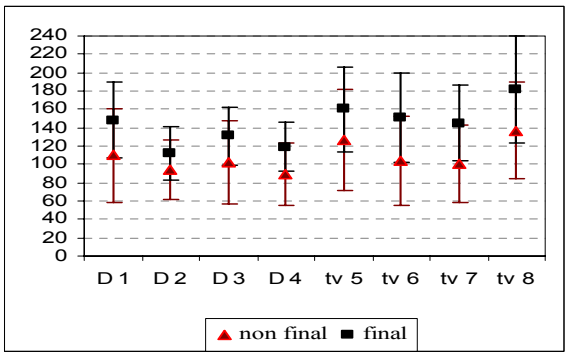

Figure 1: Mean duration (in ms. ± standard deviation) of ACCENTED nuclei (non-final and final position).

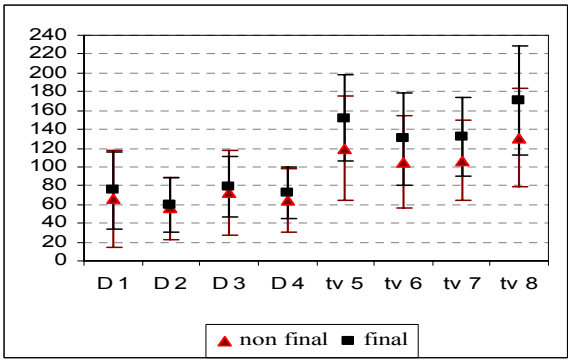

Figure 2: Mean duration (in ms. ± standard deviation) of STRESSED nuclei (non-final and final position).

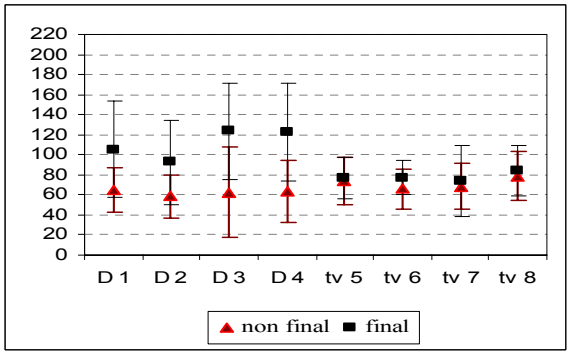

Figure 3: Mean duration (in ms. ± standard deviation) of UNSTRESSED nuclei (in non-final and final position).

3.2. Position

Figures 4 and 5 plot data for final and non final position.

Vowel durations generally increase in final positions, except for stressed syllables in dialogic speech and for unstressed syllables. The first issue has been discussed above. Unstressed syllables, representing the unmarked pole in the prominence scale, would be generally expected to present constant values, both in final or non-final position in the prosodic group. Instead, dialogic material seem to contradict such hypothesis as far as data on final position are concerned.

As shown in a previous study on dialogic speech [27], the duration of the *nucleus* of the last syllable of a prosodic group, also depends on the type of pitch movements associated with that syllable: high tones and *plateaux* entail longer durations than low tones. The phenomenon pointed out here is then strictly related to specific continuation contours typical of the dialogic interaction, which are not used in formal and monologic texts like the TV News: so that it also would be due to intonational factors.

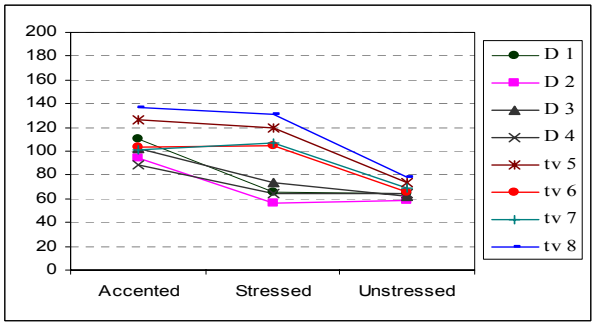

Figure 4: Mean values of accented, stressed and unstressed nuclei in NON-FINAL position.

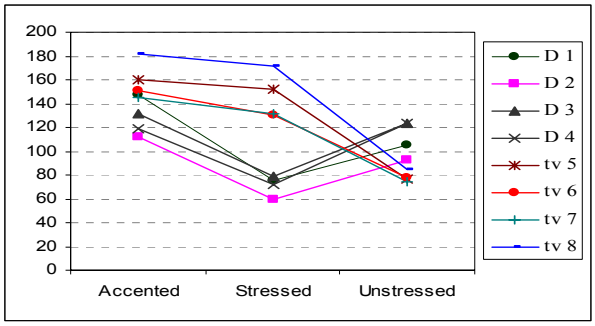

Figure 5: Mean values of accented, stressed and unstressed nuclei in FINAL position.

4. Conclusion

Duration is confirmed to be a reliable acoustic parameter related to prominence in Italian and the increase in its values can be gradually reinforced by the occurrence of an accent, mainly in case of final lengthening. Similar correlations with the intonational events seem to be valid also for unstressed final syllables. Thus, in connected speech, the temporal patterns would be someway correlated to all the layers of the prosodic structure, as for their phonetic realization.

Rhythm is directly connected to duration, as rhythmical strong syllables frequently lengthen with respect to preceding syllables; anyway, cases in which such increase does not take place suggest that perceptual effects and the interaction with other acoustic cues should be better investigated (also [37]).

Intonation seems to be active in determining temporal variation: f0 trends, particularly pitch accents and boundary tones, are frequently correlated to duration variability. Lengthening and shortenings, in fact, are also affected by relevant pitch movements and by transition zones: phonetic segmental deletion, which also show clear correspondence

with the same prosodic factors [34][35], could then represent the last step of this scale of phonetic variation.

5. References

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

[2] Nespor, M., 1993. *Fonologia*. Bologna: Il Mulino.

[3] Mancini, F.; Voghera, M., 1994. Lunghezza, tipi di sillabe e accento in italiano. *Archivio glottologico italiano* LXXIX, 51-77.

[4] Ladd, D., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[5] Hirst, D.; Di Cristo, A., 1998. A survey of intonation systems. In *Intonation Systems. A survey of twenty languages*, D. Hirst, A. Di Cristo (eds.). Cambridge: Cambridge University Press, 1-44.

[6] Gussenhoven, C., 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[7] Farnetani, E.; Kori. S., 1982. Lexical stress in spoken sentences. A study on vowel duration and formant pattern. *Quaderni del CSRF* I, 104-133.

[8] Marotta, G., 1985. *Modelli e misure ritmiche. La durata vocalica in italiano.* Bologna: Zanichelli.

[9] Vayra, M.; Fowler, C. A.; Avesani, C., 1987. Word-level coarticulation and shortening in Italian and English speech. *Studi di Grammatica Italiana* 13, 249-269.

[10] Farnetani, E.; Kori. S., 1990. Rhythmic structure in Italian noun phrases. A study on vowel durations. *Phonetica* 47, 50-65.

[11] Farnetani, E.; Zmarich, C., 1997. Prominence patterns in Italian: an analysis of F0 and duration. *ESCA Workshop on Intonation: Theory, Models and Applications*, Athens.

[12] D'Imperio, M.; Gili Fivela, B., 1997. Focus, phrasing and boundary phenomena in Italian read speech. *ESCA Workshop on Intonation: Theory, Models and Applications*, Athens, Greece, 91-97.

[13] D'Imperio, M.; Rosenthall, S., 1999. Phonetics and phonology of main stress in Italian. *Phonology*, 16, 1-28.

[14] Avesani, C.; Vayra, M.; Zmarich, C., 2007. On the articulatory bases of prominence in Italian. *Proceedings of the XVIth ICPhS*, www.icphs2007.de, 981-984.

[15] Albano Leoni, F.; Cutugno F.; Savy, R., 1995. The vowel system of Italian connected speech. *Proceedings of the XIIIth ICPhS,* (4), 396-399.

[16] Landi, R. e Savy, R., 1996. Durata vocalica, struttura sillabica e velocità d'eloquio nel parlato connesso. In *Atti del XXIV Convegno Nazionale dell'AIA*, 65-70.

[17] Albano Leoni, F.; Crocco, C.; Giordano, R., 2001. Le durate di nuclei vocalici in diversi tipi sillabici tratti da dialoghi del corpus AVIP. In *Atti del XXVIII Convegno Nazionale dell'AIA*, Trani, Bari: Addante, 233-236.

[18] Dell'Aglio, M.; Bertinetto, P. M.; Agonigi, M., 2002. Le durate dei foni vocalici in rapporto al contesto nel parlato di locutori pisani. Primi risultati. In *Atti GFS - XII*, A. Regnicoli (ed.). Roma: Il Calamo, 53-59.

[19] Greenberg, S.; Carvey H.; Hitchcock L.; Chang, S., 2003. Temporal properties of spontaneous speech. A syllable-centric perspective. *Journal of Phonetics* 31, 465-485.

[20] van der Hulst, H.; Ritter, N. A. 1999. *The syllable. Views and facts.* Mouton de Gruyter: Berlin - New York.

[21] Moore, B. C. J., 1997. Aspects of Auditory Processing related to Speech Perception, In *The Handbook of Phonetic Sciences,* W.J. Hardcastle, J. Laver (eds.). Cambridge MA: Blackwell Publishers, 539-565.

[22] Kohler, K.J., 2006. Paradigms in Experimental Prosodic Analysis: From Measurements to Function. In *Methods in Empirical Prosody Research,* S. Sudhoff *et al.* (eds.). Walter de Gruyter: Berlin - New York, 123-151.

[23] Gibbon, D., 2006. Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In *Methods in Empirical Prosody Research,* S. Sudhoff *et al.* (eds.). Walter de Gruyter: Berlin, 181-209.

[24] Beckman, M.; Edwards, J., 1990. Lengthenings and shortenings and the nature of prosodic constituency. In *Papers in Laboratory Phonology I,* J. Kingston, M. Beckman (eds.). Cambridge: CUP, 152-178.

[25] Nooteboom, S., 1997. The Prosody of Speech: Melody and Rhythm. In *The Handbook of Phonetic Sciences,* W. Hardcastle, J. Laver (eds.). Cambridge MA: Blackwell, 640-673.

[26] Giordano, R., 2006. Note sulla fonetica del ritmo dell'italiano. In *Analisi prosodica. Teorie, modelli e sistemi di annotazione*. *Atti del II Convegno Nazionale AISV*, R. Savy, C. Crocco (eds.). Torriana: EDK Editore. CD ROM. Also available online: www.parlaritaliano.it

[27] Giordano, R., (forthcoming). La posizione finale nel gruppo prosodico: una indagine sulla variabilità fonetica. In *Atti del II Convegno internazionale del Gruppo di Studio sulla Comunicazione Parlata.* E-book. Napoli: Liguori (http://www.liguori.it/).

[28] Pettorino, M.; Giannini, A., 2005. Analisi delle disfluenze e del ritmo di un dialogo romano. In *Italiano parlato. Analisi di un dialogo*, F. Albano Leoni, Giordano R. (eds.). Napoli: Liguori, 89-104.

[29] Sorianello, P., 1994. Il processo dell'allungamento prepausale: dati ed interpretazioni. *Quaderni del Dipartimento di Linguistica – Firenze* 5, 47-73.

[30] Albano Leoni, F.; Cutugno, F.; Savy, R. (eds.). CLIPS (*Corpora e Lessici di Italiano Parlato e Scritto*). Available online: http://www.clips.unina.it

[31] Turk, A.; Satsuki, N.; Sugahara, M., 2006. Acoustic Segment Durations in Prosodic Research: A Practical Guide. In *Methods in Empirical Prosody Research,* S. Sudhoff *et al.* (eds.). Walter de Gruyter: Berlin, 1-27.

[32] Zellner Keller, B., 2002. Revisiting the Status of Speech Rhythm. *Proceedings of Speech Prosody 2004,* 727-730.

[33] Lindblom, B., 1990. Explaining phonetic variations: a sketch of the H&H theory. In *Speech Production and Speech Modeling*, W.J. Hardcastle, A. Marchal (eds.). Dordrecht: Kluver, 403-439.

[34] Savy, R.; Cutugno, F. 1997. Hypospeech, vowel reduction, centralization: how do they interact in diaphasic variations? In *Actes du XVIe Congrès International des Linguistes*, B. Caron (ed.). Oxford: Pergamon. ID 0185

[35] Savy, R. 2001. L'interfaccia tra livelli di analisi del parlato: rapporti tra riduzioni segmentali e schemi prosodici. In *Atti del XXXIII Congresso Internazionale della Società di Linguistica Italiana*, F. Albano Leoni *et al.* (eds.). Roma: Bulzoni, 309-328.

[36] Kohler, K.J., 2006. What is Emphasis and How is it Coded? In *Proceedings of Speech Prosody 2006*, R. Hoffmann, H. Mixdorff (eds.). Dresden: TUD press.

[37] Alfano, I; Llisterri, J.; Savy, R., 2007. The perception of Italian and Spanish lexical stress. *Proceedings of the XVIth ICPhS*, www.icphs2007.de, 1796-1799.