CORRELATION BETWEEN SEGMENTAL REDUCTION AND PROSODIC FEATURES IN SPONTANEOUS SPEECH: THE ROLE OF *TEMPO*.

Francesco Cutugno and Renata Savy
CIRASS - University of Naples "Federico II"
Italy

ABSTRACT

This work aims to point out some relations between segmental reduction and suprasegmental structures in Italian.
Following a previous study that observed an increase of reduction probability to occur after a pitch or nuclear accent in the tone-unit, we tried to verify the existence of relations between the amount of reduction and speech-rate variation in pre-accent and post-accent positions, introducing the notion of local speech rate (l-SR) as the pattern of the inverse values of syllable duration.
In a first stage, we classified three different l-SR patterns in the considered positions. In a second stage, we calculated the percentage of reductions occurring in each part of the scheme and the relation with the l-SR patterns.
Final results show that, although a coincidence of greater amount of reduction in post-accent positions with increasing tendency of l-SR does exist, it is not systematic; l-SR decrease is not able to inhibit reduction.

1. INTRODUCTION

Aim of this paper is to study the relations between segmental reduction and suprasegmental structures in spontaneous speech.
It is well-known that segmental reduction phenomena can be classified within the general framework of the natural tendency to hypoarticulation of speech. Hypoarticulation affects the whole segmental structure of speech; during production most of the articulatory target positions are not reached, consequently acoustic parameters corresponding to the gestures appear to be significantly different from the ones observed in the formal production. This phenomenon finds its best description in the 'target undershoot' model [4, 5, 8, 9].
Within this framework particular attention has been given to vowel and syllabic reduction phenomena in many languages different from Italian. Van Bergem [8], for example, finds important relations between acoustic vowel reduction (defined as "a gradual shift of the acoustic features of vowel away from their "target" features towards those of the schwa"; [p. 6]) and various linguistic, sociolinguistic and phonetic factors. Among these factors we consider the ones connected to the prosodic structure as particularly relevant:

-*word stress*: syllables carrying the lexical stress are usually pronounced with higher accuracy than the unstressed ones; nuclear vowels are, in the first case, close to the ideal target;
-*sentence accent*: although this is not a stable property of words, being assigned to them under particular conditions related to the intonational structure of the string, vowels in accented syllables are less reduced than ones in unaccented syllables;
-*tempo* (or equivalently *speech rate*): seen as a function of speech style, it is only in part responsible for the presence of reduction and shortening of vowels, mainly if it is associated with a decrease of accuracy in the pronunciation.

As far as studies about Italian are concerned, whereas some of them confute this hypothesis [2] more recent papers support the fact that Italian, like many other languages, presents vowel reduction phenomena [1, 3, 10] which mainly appear in spontaneous speech. Albano Leoni et al. [1], Savy & Cutugno [6] showed the strict connection existing between word stress and vowel reduction: unstressed and word-final vowels show an high degree of reduction if compared to stressed ones, especially (but not uniquely) in spontaneous speech.
The present paper is based on a previous study on segmental reduction phenomena affecting Italian vowels and syllables in word-final position [7]. These phenomena have particular relevance on the morphological level as the above units are realisations of nominal and verbal inflectional suffixes. In [7] Savy studied the following types of reductions: 1) changes in vowel quality consisting either in the centralisation toward the *schwa* target, or in a timbre substitution (R); 2) total deletion of the nuclear vowel or of the whole syllable (D). She observed that reductions can be related to presence of sentence accents within intonational profile (f_0 patterns). Her results show a relation between the place in which most of the observed segmental reductions occur and the position of the accents: within the tone unit (TU), syllables immediately following a pitch or a nuclear accent have higher probability to present reduction, especially when they fall across the boundaries of intonational units like TU or intermediate phrases.
On this basis, Savy proposed a model stating that the portion of intonational profile preceding a pitch accent is protected against reduction because of the relevance it plays for prosodic processing. At the same time, during the f_0 transition after the accent, "environmental" conditions for segmental reduction arise. In this view, phases containing a sort of "prosodic tension" alternate with more "relaxed" (and favouring reductions) ones.
In this paper we will present results of a further analysis on the same material based on the observation of speech rate distributions within the TU, just in the same portion of signal that were previously defined as tense and relaxed.

Our aim is to verify the existence and the nature of relations between speaking rate variations within the TU in the above cited portions and the number of segmental reductions observed in our corpus.

2. METHODS

The present work is a pilot study conducted on a corpus consisting of a set of 18 TU. Speech material has been selected from the wider corpus used in [7] because of the systematic and relevant presence of segmental reductions in it. This material is derived by recordings of a set of spontaneous conversations. During the dialogues both tape recorder and microphones were hidden, so that all speakers were unaware of being recorded.
All the used TU have been preliminarily divided into syllables on the basis of Italian phonological rules, a second review of this division has then been performed according to the effective phonetic realisation.
Recordings have been digitised and spectroacoustically analysed. f_0 curves were drawn and the position of pitch accents (pa) and nuclear sentence accents (PA) have been determined using both this patterns and auditory analysis of the TU. Syllables defined according to previous description have been segmented and their duration was measured. As already stated before, reduction can extend its effect until the limit of the total deletion of some syllables. In all cases in which acoustic analysis did not allow us to individuate any evidence of the presence of a phonologically expected syllable we marked it as absent and assigned it a duration equal to zero.
On the basis of these duration measurements, we introduced the notion of *local speech rate* (l-SR) defined as the pattern of the inverse values of syllable duration. Each of this values expresses then the syllable/sec rate as a local value within the speech chain, while the sequence of these values is a discrete evaluation of the mathematical derivative of the speech waveform expressing (velocity)/speech rate as a function of time. A further simplification can be included in this representation if we consider the syllable sequence as an independent variable on the x-axis instead of time as is shown in figs. 1 a) and b), which both give examples of the shape that these curves can assume.

Text:

TU14: In queste due lingue così...diverse
In these two languages so...different

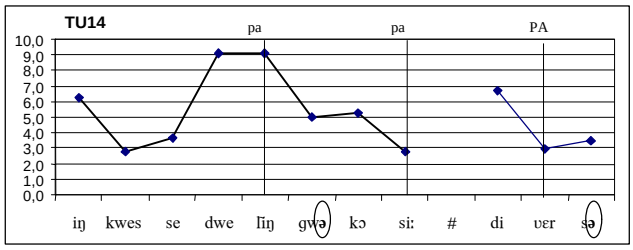

Figure 1a: example of speech rate chart.

Syllable sequence is represented on x-axis while speech rate is expressed in terms of syll/sec on the y-axis. Vertical dotted lines, marked as 'pa' and 'PA' indicate respectively the position of an intonational pitch accent and of a nuclear accent. Circled vowels are affected by reduction.

Text:

TU28: Come si chiama il latino, lo sa?
how PRON-III-SING-REFL call the latin, it know
What is the name of Latin, do you know it?

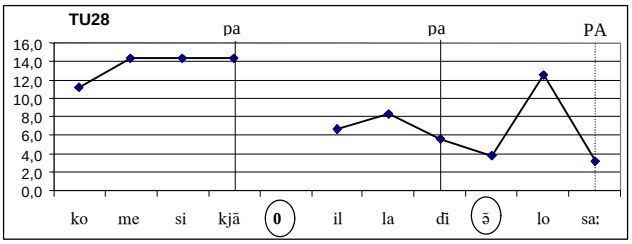

Figure 1b: another example showing a different velocity pattern.

TU showed in fig. 1a is characterised by the presence of three 'pa' (syllables [liŋ], [si] and [vɛr]), while the last accent is the nuclear one (PA) and it corresponds to a *focus*. The symbol # indicates a pause.
TU showed in fig. 1b presents a syllable deletion after the first 'pa', in order to avoid that the curve diverges we simply omit the point in the curve indicating the missing syllable in the x-axis. Implicit in this model is then present the idea that speech rate should be calculated not on the number of produced syllables per time unit but *on the number of intended syllables.*

We have calculated the total amount of positions with respect to the accents in our 18 TU; results are shown in table 1:

Total positions	112
pre-accent positions	55
post-accent positions	57

Table 1: number of positions calculated

For each of these positions, when possible, we may analyse speech rate in the groups of syllables preceding (pre-) and succeeding (post-) accents. In particular we define a pre-accent syllable group as the sequence including the last unit before the accent: in the example in fig 1a the first pre-pa group is formed by syllables [iŋ] [kwes] [se] [dwe]. At the same time we define post-accent syllable group as the sequence including the unit carrying the accent and extending until the last syllable of the

lexical unit: in the example in fig. 1b the two syllables [ti][no] define the post-pa group.

3. RESULTS

A first result obtained in this work consists of a classification of the speech rate (l-SR) patterns within TU with reference to the scheme provided in [7] which assigns to pitch and nuclear accents (pa/PA) the role of key points discriminating between "tense" and "relaxed" zones in the stream of speech. Our classification leads to three different patterns: SR increasing, SR decreasing, l-SR constant.
Figures 2a and b illustrate how, on the base of this classification, examples reported in fig. 1a and b can be stylised and categorised. Obviously the y-axis expresses now an arbitrary unit as we are now simply describing a general tendency. As far as the x-axis is concerned the alignment between phonetic transcription and time events is only indicative.

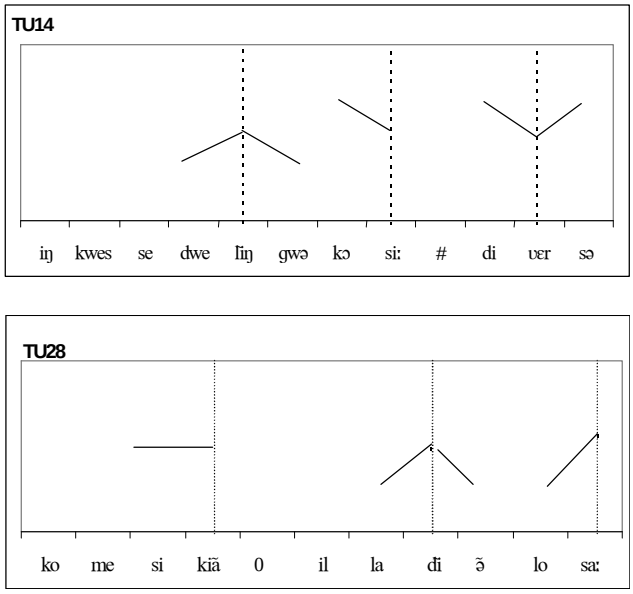

Figure 2 a and b: Speech rate stylisation for same TUs as in fig. 1a and b.

In some cases the stylised l-SR pattern cannot be defined. It happens when a syllable is totally deleted (like in 1b/2b), or when a pause (#) occurs (like in 1a/2a), or in cases in which accent coincide with the first or the last syllable in the phrase (like the last syllable in 1b/2b).
We then counted the amount of pre- and post- accent position that fall into the l-SR patterns above classified. Table 2 shows results.

	pre-accent positions	post-accent positions
SR increase	33%	**65%**
SR decrease	40%	15%
SR constant	19%	5%
Total on all position	*92%*	*85%*

Table 2: amount of pre- and post- accent position vs SR pattern classification.

Table 2 indicates that, although in pre-accent position l-SR does not show any regular tendency, l-SR seems to clearly increase in post-accent positions as indicated by the value in bold in the table. This means that in the 'relaxed' phases immediately following a phrase accent an acceleration inducing the shortening of unstressed syllables can be frequently observed in spontaneous Italian.
Is this result in some relation with the presence of segmental reduction? Or, in other words, does a correspondence exist, based on prosodic factors, between the segmental reduction and the shortening of speech sounds?
In order to answer these questions with calculated the percentage of reductions occurring in the tense and relaxed positions available in our corpus, obtaining the result shown in table 3.

	tense phase (pre-)	relaxed phase (post-)
R-distribution/total R (%)	10,6%	89,4%
R-distribution/total of positions with R (%)	9,1%	73,7%

Table3: Reduction distributions in tense and relaxed phases.

In the first row of this table the global amount of reductions is divided according to their position in the phrase, in the second row the same data are considered with respect to the number of pre- and post- accent positions presenting a reduction.
These data are coherent with the results already reached in [7]: about 90% of reductions fall in post-accent positions (moreover the whole set of deleted syllables occurs in this position).
At a first glance, this result seems to confirm that reductions and increase of SR are connected. But this assertion is weakened if we look at the data of table 4:

	pre-acc. pos.	post-acc. pos.
l-SR increase	6,3%	62%
l-SR decrease	13,6%	56%
l-SR constant	0%	66%

Table 4: Presence of reductions compared to l-SR tendency.

Reductions are then less present in all tense phases, where, however, percentage slightly grows when l-SR decreases. In relaxed phases the different behaviours of l-SR seems not to interfere with R distributions. The maximum value of the second column in table 4 is 66% obtained when l-SR remains constant, but this value is scarcely representative of the real tendency as l-SR constant group is under-represented (these cases represent only 5% of the total amount of observed positions).

4. DISCUSSION

Aim of the present paper was to verify the existence of correlation between temporal structure and segmental reduction in spontaneous speech. We started from the assumption that the latter occurs mainly in well defined portions of the speech stream related to shape of the intonational pattern.

Our results agree with the ones obtained by van Bergem [8], as we showed that, although this correlation exists, it cannot be considered as systematic. An increase in l-SR (and the related decrease in syllable length) can create the condition for the occurrence of reduction, nevertheless a l-SR decrease or a constant shape cannot invert the tendency and is not able to inhibit reduction. At the same time during the prosodically "tense" phases speech rate may increase without causing segmental reduction. As already suggested by van Bergem [8] these result partially disagree with the Lindblom model [4] that sees the "target undershoot" strictly related to segmental duration of speech sounds. On the contrary, we showed that it is possible that, in the speech chain, both short segments accurately produced and long hypoarticulated segments can be observed. This can depend on speaking style, as suggested in [8], but also on the position into which segments fall with respect to the intonational pattern.

In substance, we could conclude that *Tempo*, as previously defined, plays then a role as a 'supporting' factor, but in a subordinate measure to the 'planning-modulation-declination' scheme of intonational pattern.

We think that only when f_0 does not present any significant movement, reduction can directly occur depending on the speech rate increase: we need to demonstrate this last assumption also if some first evidence supporting it has been already found in [7].

Furthermore, we need to perform more investigations on the relations between reduction of type D, i.e. the syllabic total deletion, and l-SR.

In conclusion we believe that relations between segmental reduction and prosodic structures in Italian spontaneous speech must be approached also on the base of an accurate analysis of the rhythmical structure of the speech-stream.

REFERENCES

[1] Albano Leoni, F., Cutugno, F. and Savy, R. 1995. The vowel system of italian connected speech. In: *Proceedings of XIIIth ICPhS*, Stockholm., **4**, 396-99.

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

[3] Giannini, A. and Pettorino, M. 1992. *La fonetica sperimentale*. Napoli: Ed. Scientifiche Italiane.

[4] Lindblom, B.E.F. 1963. Spectrografic study on vowel reduction , *JASA*, **35**, 1773-1781.

[5] Lindblom, B.E.F. and Moon, S.J. 1988. Formant undershoot in clear and citation-form speech, *Phonetic Experimental Research Institute of Linguistics University of Stockholm (PERILUS*) **8**, 21-33.

[6] Savy, R. and Cutugno, F. 1998. Hypospeech, vowel reduction, centralization: how do they interact in diaphasic variations? *Proceedings of the XVIth International Congress of Linguists*, in press.

[7] Savy, R. 1999. Riduzioni foniche nel parlato spontaneo: il ruolo della morfologia nell'interpretazione del messaggio e nella comunicazione, Unpublished PhD Thesis, University of 'Roma Tre'.

[8] van Bergem D. 1995. *Acoustic and lexical vowel reduction*, Amsterdam: IFOTT.

[9] van Son, R. 1993. *Spectro-temporal features of vowel segments*, Amsterdam: IFOTT.

[10] Vayra, M. 1989. Aspetti della riduzione vocalica in italiano. In *Atti del XVII Conv. Naz. AIA*, Parma, 421-426.