

QUANTIFYING THE RANGE OF VOWEL REDUCTION IN ITALIAN

Edda Farnetani* and M.Grazia Busà†

**Istituto di Fonetica del CNR, Padova, Italy*; †*Università di Bologna, Italy*

ABSTRACT

New phonological accounts view vowel deletion as the endpoint of a gradient articulatory reduction process rather than the outcome of phonological rules. Languages considered as belonging to different rhythmic types can be seen as sharing the same spatiotemporal reduction process, though at different degrees and frequencies, in a continuum from full to deleted vowels. Our study was designed to locate Italian along this continuum. The speech material was collected to test the main factors which have been reported to favor or inhibit reduction/deletion in the languages where it has been studied. The results indicate that in Italian, cases of apparent deletion are limited to 2%, and are restricted to specific vowels or consonantal contexts. For both deletion and durational reduction, the favoring factors are stress, vowel quality, consonant context, and speech rate, which confirms that they are part of the same reduction process. The data suggest that Italian is a non-reducing language.

1. INTRODUCTION

The present work is part of a larger study designed to assess the differences in vowel reduction and deletion between Italian speakers and English learners of Italian in the production of Italian utterances. This paper presents the results for Italian speakers, which allow us to trace, in a preliminary way, some differences between Italian, a syllable-based language, and other languages with similar and different rhythmic structures. Relevant features of a syllable-based language should be, first, the preservation of all its syllables also in rapid connected speech (by avoiding vowel deletion as much as possible); and, second, the preservation of the identity of reduced vowels (by avoiding extreme centralization and neutralization to schwa). The structure of the Italian lexicon speaks in favor of this hypothesis. The most frequent and most commonly used content words are bi- and tri-syllabic; the lexical inventory is rich of minimal-pair words differing solely for the presence vs absence of an unstressed syllable - e.g. /'libro/ "book" vs /'libero/ "free"; a large number of trisyllables differ only for the quality of the unstressed vowel - e.g. /pa'lato/ "palate" vs /pi'lato/ (proper noun) vs /pe'lato/ "bald". If the lexicon complies with the syllable-based-language proposition, and if the spoken language complies at least in part with the phonological contrasts expressed in its lexicon, then this proposition should be confirmed. Phonological descriptions indicate that in Italian there is no vowel reduction or vowel deletion, but there are no acoustic studies to confirm or disconfirm this strong claim. In the light of the two hypotheses outlined above, our study intends to assess the frequency of occurrence of vowel deletion, and the degree of reduction in duration and in formant structure for undeleted vowels.

As for deletion, the literature on other languages indicates that in the stressed-timed languages English [1, 2] and German [3] the deletion of weak vowels is quite extensive and ranges across a variety of contexts. It is important to note that in the theoretical account of gestural phonology [4], vowel deletion is

not viewed as the outcome of a phonological deletion rule, but as

the endpoint of a continuous reduction process. Therefore deletion is itself gradual, and nothing more than reduction at its highest degree: the apparently deleted vowel is still there, but is hidden more or less extensively by the neighboring consonantal gestures. This novel view of vowel deletion is supported by the acoustic studies on the topic, which (by means of a fine-grained analysis) have all found acoustic traces of the presence of the 'deleted' vowel. So has the English study on vowel deletion and contrast neutralization of paired-words like 'sport' and 'support' carried out by Manuel et al. [5]. As for languages other than English, Dauer's study of Greek high vowels [6] shows that they are subject to a high degree of reduction in unstressed syllables, to the point of being completely devoiced. This occurs especially in post-tonic position and in a voiceless environment, particularly in /s/ environment. Mary Beckman's [7] ample overview of vowel deletion studies shows that in Montreal French, high vowels are reduced at different degrees from 'mere reduction' to 'devoicing', to 'syncope', especially when preceded by a voiced or voiceless fricative. Similarly, the devoicing of high vowels in Japanese (object of numerous studies) ranges from 'partially devoiced' to 'fully devoiced' or 'deleted'.

Altogether the auditory and acoustic studies conducted up to now indicate that extreme vowel reduction can occur in any language and that the process is similar across languages of different rhythmic structure. It appears, however, that in stress-timed languages deletion is invasive, occurring in high vowels, when unstressed [2], and in all the unstressed syllables with an underlying schwa vowel [7], whilst in syllable-timed and mora-timed languages it seems to be restricted to specific vowels in specific consonantal environments, although in Montreal French it appears to be more extensive than in Greek and Japanese.

The present study has taken into account, besides two well known factors associated with reduction in connected speech - speaking rate and prosodic position of the word - also the factors that seem to favor/disfavor extreme reduction in other languages: consonant context and vowel quality (which seem to be primary factors in Greek and Japanese). Other important factors outlined by Beckman [7] are: phonotactic constraints on the shape of a syllable (extreme reduction may be favored or disfavored, depending on the type of cluster, legal or illegal, resulting from it); position of the unstressed syllable with respect to the stressed syllable (post-tonic vowels are likely to be more often deleted than pre-tonic); constraints associated with neutralization of a lexical contrast, i.e. extreme reduction might be disfavored in those words where the possibility of reanalysis would lead to neutralization of a lexical contrast. From the data available in other languages, it is not clear whether this perceptual constraint really works. In the present study, many trisyllables have a corresponding bisyllabic word differing only for the absence of the key unstressed syllable, hence, this constraint too can be tested.

2. METHODS

2.1. Speech Material.

The phonetic material consists of 18 meaningful trisyllabic words with prosodic structures /+ - -/ and /- + -/. Nine of these trisyllables have paired bisyllables differing only for the absence of the key vowel (e.g. /'alito/ vs /'alto/; /se'para/ vs /'spara/). The other nine trisyllables do not have a paired bisyllable in Italian, either because the deletion of the unstressed syllable would yield an illegal cluster (e.g. /'lutSido >*lutSdo/) or because the paired bisyllable does not exist in the Italian lexicon. In the present work a quadrisyllabic word was also used to better balance the number of unstressed vowels of different quality. Target vowels were the seven Italian vowels /i,e,ɛ,a,ɔ,o,u/ occurring in stressed syllables and the five /i,e,a,o,u/ occurring in unstressed syllables. The unstressed vowels were either in pre-tonic word-initial position, or in post-tonic word-medial position. The consonants preceding and following the unstressed vowels were grouped into the seven following sequences: affricate+stop (indicating prevocalic affricate and postvocalic stop); fricative+stop, fricative+/t/, stop+stop, stop+/r/, /l/+stop, and /n/+stop. Stop consonants were either voiced (/b, d/) or voiceless (/p, t, k/), fricatives and affricates were all voiceless (/f, s/ and /tʃ/).

The words were embedded in phrase-non-final position of two short-stories texts.

Four speakers of standard Northern Italian, one male (GA) and three females (BB, EF, and LD), read the stories at a comfortable rate (Normal Rate) for five times and at a fast rate for two times. The digital recording took place in a sound-treated room. Using the Multispeech software, the speech material was downsampled at 14.7KHz, and spectrographic analysis was performed using a 125-point FFT, Hanning window, and a frame advance of 5 ms.

It must be remarked that the choice of real, commonly used words, and of specific CVC sequences where the vowels are expected to be deleted in other languages, has led to a rather unbalanced speech material where the vowels /i/ and /e/ occur more frequently than the other vowels and where not all the vowels share the same consonantal context. Therefore the present results are to be considered preliminary.

2.2. Analysis

Duration, F1 and F2 were measured in all test vowels. The vowel boundary was set at the point of disappearance of the second formant. We considered extremely reduced the vowels with very short durations and no evidence of voiced formants. These vowels were excluded from statistical analysis, and are discussed apart in 3.3. Measures of F1 and F2 were taken at vowel midpoint over three consecutive frames whose values were averaged. The measurements were all checked by visual inspection of formant tracks overlaid on spectrograms. We also measured the total word duration.

Statistical analyses were carried out for each subject separately. As for DURATION, sets of Anova's were run to test the effects of the following variables: STRESS, RATE, VOWEL QUALITY for the five vowels /i, e, a, o, u/ in stressed and unstressed syllables. For the unstressed vowels, we tested their position in

relation to stress, i.e., PRE-TONIC vs POST-TONIC and CONSONANT CONTEXT. The significance level was set at 97.5%.

3. RESULTS

3.1. Vowel duration

With this analysis it was possible to assess the degree of temporal vowel reduction in Italian, and to have an idea of which vowels, in what contexts, might be the most probable candidate for deletion. Factor STRESS resulted highly significant for all the subjects ($p=0.000$) with duration ranging from 77.24 ms for stressed Vs to 45.19 ms for unstressed Vs (data averaged across subjects). Globally, the duration of a vowel, when unstressed, is reduced by 40.6% of its duration. Effects of position of the unstressed vowel with respect to stress have a similar trend across the subjects, and indicate that post-tonic vowels tend to be shorter than pre-tonic ones; the differences reach the significance level only in two subjects, BB (54.64 ms for pre-tonic V vs 45.19 ms for post-tonic V; $p=0.001$) and GA (52.50 ms vs 40.49 ms; $p=0.000$). It is worth noting that for these two speakers the average V durations and word durations are longer than for the other two subjects, thus it is possible that in slower speakers minor durational differences show up better than in faster speakers.

The effect of SPEECH RATE is also highly significant for all the subjects ($p=0.000$). The magnitude of durational changes as a function of rate is smaller than that associated with stress: high rate reduces a stressed V duration by an average 20.50% and an unstressed V by an average of 15.10%. So, as rate increases, stressed and unstressed vowels tend to become more similar in duration.

Another significant factor is VOWEL QUALITY. It is well known that segments have their own intrinsic duration associated with the requirements for their production. Previous studies on Italian show that /a/ is always longer than /i/ [8] but little is known on the other three vowels. In the present study there are a few discrepancies among the subjects regarding the ordering of vowels from the longest to the shortest. Globally, the data indicate that the longest vowels are /o/ and/or /a/, followed by /e/, followed by /u/ and /i/. The values averaged across subjects are: /o/ = 73.40 ms, /a/ = 68.78 ms, /e/ = 59.44 ms, /u/ = 50.65 ms, and /i/ = 49.83 ms.

CONSONANT CONTEXT. As mentioned above, with the present speech material it is not possible to have all consonantal contexts for each of the five vowels. However, vowels /i/ and /e/, which occur most frequently in our corpus, have a consonantal environment sufficiently differentiated to permit a reliable account of contextual effects. Since the two vowels differ in intrinsic duration, the effects of the flanking consonants were analyzed separately for each vowel. The global results indicate that the consonantal context has significant effects on /i/ duration for BB, GA and LD (p values ranging from 0.000 to 0.007). For subject EF the trend is the same but significant differences occur only between affricate+/d/ vs affricate+/t/ (35.40 ms vs 26.83 ms). Altogether the data indicate that the duration of /i/ decreases from /l+/t/ and /p+/t/ environments (average values: 45.37 ms, and 41.02 ms, respectively), to fricative+/k/ (33.46 ms) to affricate+/d/ (32.17 ms), to

affricate+/t/ (24.79 ms). For /e/ the contextual effects are significant for all the subjects (p values ranging from 0.000 to 0.003), and indicate that /e/ duration decreases from fricative+/r/ and /n+/t/ environments (average values: 60.71 ms and 57.61 ms respectively) to /b+/r/ and /p+/r/ (46.76 ms and 40.71 ms respectively) to fricative+/p/ environment (38.42 ms). Notice that the durations of /i/ and /e/ in the context of fricative+stop are similar, whilst in the context of fricative+/r/, vowel /e/ is much longer. In the two former examples the fricative is sibilant /s/, while in the fricative+/r/ environment the fricative is labiodental /f/. The quite long /e/ duration in /f+/r/ context can be related to the different constraints associated with the production of /s/ and /f/, and possibly to the lengthening effect of the following monovibrant /r/.

The data clearly indicate that for both /i/ and /e/, prevocalic voiceless affricates and sibilants, combined with postvocalic voiceless stops, have the strongest reduction effect on vowel durations. The effects of prevocalic stops are much weaker than those of sibilants or affricates (notice the difference in /i/ duration between the environment of /p+/t/ and that of affricate+ /t/ - 41.02 ms vs 24.79 ms). Finally, it appears that environments of nasal or lateral +stop have the least reduction effect.

Consonant voicing also has significant effects on vowel duration: as can be seen in the data for /i/, the vowel tends to be shorter in the affricate+/t/ than in the affricate+/d/ context.

3.2. 'Deleted' Vowels

On a total of over 560 productions of unstressed vowels, only twelve cases of deletions were observed, a little more than 2% of the total. Eleven deletions occurred for /i/ (7% of /i/ productions), one case was observed for /e/. These cases of extreme reduction are well balanced across three of the subjects. V deletion occurred only once for subject BB, possibly because she spoke at a slower rate than the others (see 3.1.). The majority of deletions occurred at a fast speech rate (8 over 12). The context is restricted to a few consonants. Of the twelve reductions, seven occurred in the context of affricate+voiceless stop, (/ˈtatSito/ - "tacit"), two in the affricate+voiced stop context (/ˈlutSido/ - "shiny"), one in the context of sibilant+/k/ (/siˈkuro/ - "confident"), and one in /p+/t/ context (/ˈkapita/ - "it happens"); the one involving /e/ occurred in the context of sibilant+/p/ (/seˈpara/ - "it divides"). Considering vowel type and context, deletion of /i/ in voiceless affricate+stop context scored 25%. These patterns faithfully reflect those of durational reductions, confirming that the two are only different expressions of the same reduction process.

The data clearly indicate that the context is the factor mainly responsible for vowel deletion, that speech rate is another important factor, and that the intrinsically short high vowels are the best candidates. It appears that deletion involves both post-tonic and pre-tonic syllables. The hypothesis that deletion would be disfavored when reanalysis neutralizes lexical contrast is not supported by the present data where three items with a deleted vowel have paired bisyllables in the text.

3.3. The deletion process

The cases of vowel deletion were analyzed in detail by means of

spectrograms, power spectra, and intensity curves. As in other studies of vowel deletion, a distinction can be made, in a voiceless consonantal context, between partially devoiced and completely devoiced vowels (see [7]). There are cases of both types in this study.

3.3.1. Complete devoicing. Complete devoicing of the key unstressed vowels occurred more often than partial devoicing in the context of /tS/+stop. In this context, traces of a completely devoiced /i/ were detectable from the location of a consistent peak of energy at a frequency compatible with the second formant of /i/ (around 2.200Hz). This configuration occurred towards the end of the friction noise. The hidden vowel could be distinguished from the affricate owing to the difference in the frequency of the lower energy peak, which, for the affricate was usually over 3Kz. It was more difficult to detect reliable energy peaks at frequencies compatible with the first formant of /i/.

Also in the word /seˈpara/ spoken at fast rate by subject BB, the pretonic /e/ is completely devoiced. The waveform shows a long interval of friction noise followed by the silence for /p/ closure. The comparison with another fast rate version of the word with no deletions shows that in the deleted version the ratio of /s/ duration to /p/ closure duration is 1.97, whilst in the undeleted version it is only 1.15, and that the temporal interval from /s/ onset to /p/ release is the same in the two versions (238 ms and 232 ms). This means that the vowel is still there, but is completely hidden by the oral and the glottal /s/ gestures. The comparison also indicates that in some cases duration alone is an excellent indicator of the presence of a hidden vowel: consonant /s/ in this word sounds like syllabic [s] in English.

A complete devoicing of the post-tonic vowel /i/ occurred in the fast rate production of item /ˈkapita/ (subj. EF): here the presence of the vowel is inferable from the release of the first stop into a quite audible voiceless interval of 32 ms. The unvoiced release clearly indicates a blending together of the glottal abduction gestures of the two consonants across the vowel [9].

3.3.2. Partial devoicing. In the affricate+/t/ environment there are also cases of partial devoicing. And here too, towards the end of the voiceless fricated period of the preceding affricate, a consistent energy peak is in a frequency region compatible with the second formant of /i/, whilst when two or three weak vocal pulses appear, the /i/ like configuration fades away very quickly. These patterns indicate that the oral tract has started to take the configuration of the vowel when the glottis is still abducted and/or the vocal tract is still too narrow for the initiation of voicing. When the vocal folds start to vibrate the vocal tract gestures seem to be already anticipating the oral closure for the following stop. This pattern seems to result from the slow and delayed offset of the consonant gesture, which can hide more or less completely the vowel configuration. Similar patterns can also be observed in the context of affricate+voiced stop, illustrated in Fig.1.

It can be seen that the only peak of energy (around 150 Hz) is the voice bar of consonant /d/. After this point in time, the energy at high frequencies starts to decrease consistently (see

the two unmarked spectra) and fades out between the third and the fourth vocal pulse, when the only observable peak of energy is the low frequency voice bar of consonant /d/.

These data indicate that the various degrees of reduction depend on the degree of gestural blending, and suggest that the devoicing of the oral gestures of vowels can result from a

temporal delay in the adduction of vocal folds with respect to the oral gesture for the vowel. It is possible that small differences in the inter-articulator timing between the oral and the glottal gestures may be sufficient to delete a short vowel especially in a fast speech condition.

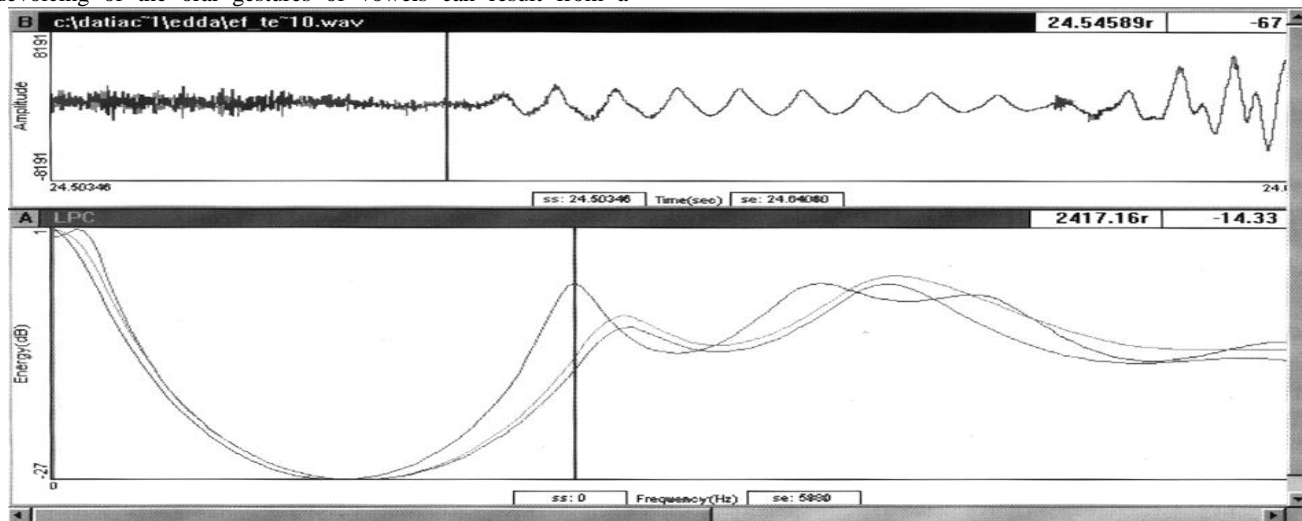


Fig. 1. Waveform of the temporal interval between the release of the /tS/ constriction and the onset of the final vowel /o/ in /lutSido/. The marked spectrum is taken at the cursor, before vocal fold vibrations are visible. The unmarked spectra were taken at the first and the second vocal pulses.

4. DISCUSSION AND CONCLUSIONS

In the accounts which distinguish languages between stress-timed, syllable-timed, and mora-timed, one of the features characterizing the first versus the second rhythmic type is the presence vs absence of vowel reduction. Italian, being one of the languages defined as syllable-timed, should show no or limited instances of vowel reduction. Non-experimental work on Italian has also claimed that Italian does not have vowel reduction at the phonological level. The newer accounts proposed by gestural phonology [4] view vowel spatiotemporal reduction and apparent cases of vowel deletion as the result of a continuous process presenting itself, at different degrees and with different frequency of occurrence, in all languages. Real vowel deletion is seen as the consequence of an extreme and habitual gestural overlap which gets reinterpreted through phonological reanalysis [7]. Our study was designed to test the possibility and extent of spatiotemporal vowel reduction in Italian.

The data, though still preliminary, are enlightening as to the degree and effects of the occurrence of the vowel reduction process in Italian. If we were to place Italian along a continuum from strong to minimal consequences on the prosodic-syllabic structure of the language, Italian would be among the languages which are only minimally affected by the reduction process. This is evidenced by the two main findings of this study, that is, that both stressed and unstressed vowels get preserved even when reduced in duration (the only effect of fast rate being a decrease in the durational differences between stressed and unstressed vowels); and that the only cases of extreme reduction/apparent deletion seem to occur independently of the possibility of lexical reanalysis. This shows that in Italian the process of vowel

reduction is not extreme, and does not effect the syllable count in a word. However the Italian rhythmic type gets defined, it is far from one like English, which allows for more dramatic reduction phenomena.

REFERENCES

- [1] Roach, P. 1991. *English Phonetics and Phonology: A Practical Course*, (2nd ed.). Cambridge: Cambridge University Press.
- [2] Gimson, A.C. 1989. *An Introduction to the Pronunciation of English*. (4th ed.). London: Edward Arnold.
- [3] Kohler, K. 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle, W.J and A. Marchal (eds.), *Speech Production and Speech Modelling*. Amsterdam: Kluwer Academic Publishers. 69-92.
- [4] Browman, C.P. and L. Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In: Kingston, J. and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press. 341-376.
- [5] Manuel, S.Y., Shattuch-Hufnagel, S., Huffman, M., Stevens, K.N., Carlson, R., and S. Hunnicut. 1992. Studies on vowel and consonant reduction. *Proceedings of the 1992 International Conference of Spoken Language Processing*, Vol. 2, 943-946.
- [6] Dauer, R.M. 1980. The reduction of unstressed high vowels in Modern Greek. *Journal of the International Phonetic Association*, 10, 17-27.
- [7] Beckman, M.E. 1996. When a syllable is not a syllable? In Otake, T. and A. Cutler (eds.), *Phonological Structure and Language Processing, Cross-Linguistic Studies*. Berlin: Mouton de Gruyter. 95-123.
- [8] Farnetani, E., and D. Recasens. 1993. Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech*, 36, 279-302.
- [9] Munhall, K., and A. Löfqvist. 1992. Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics*, 20, 111-126.