POSITIONAL EFFECTS ON STRESSED VOWEL DURATION IN STANDARD ITALIAN

Jan van Santen† and Mariapaola D'Imperio‡
† *Bell Labs – Lucent Technologies, USA*, ‡ *Department of Linguistics, The Ohio State University, USA*

ABSTRACT

We investigate the effects of intra-word position on the duration of stressed vowels in open syllables in Italian. While traditional accounts claim that vowels in all non-final position are longer than vowels in word-final position [9], we find that only penultimate stressed vowels are longer, confirming the findings reported in [2]. In addition, we show that the duration values of final stressed vowels depend critically on measurement criteria. In word-final position, stressed vowels often present a heavily glottalized region, which can be included or not in the vowel. When included, the resulting vowel duration is similar to that of penultimate vowels. We also found that penultimate vowels in three-syllable words are longer than either initial or final vowels, which is different from what has been found for stressed vowels in other languages. These results have implications for multilingual text-to-speech synthesis, in particular for the importance of methodologies that do justice to language-unique prosodic patterns.

1. INTRODUCTION

It is a well-know fact of Italian phonology that vowels in open syllables are long under stress [9, 8]. For instance, the vowel [e] of *séta* "silk" is generally long, while the same vowel in the word *sétta* "sect" is short. A derivational rule that lengthens stressed vowels in open syllables is presented in [9]; however, vowels in word-final position are excluded from this rule. No further restriction to the scope of the rule is offered regarding non-final vowels. This therefore predicts that all non-final stressed vowels are equally long, and are longer than vowels in word-final position. For example, the duration of [e] in *séta* should be the same as the duration of [e] in *sétola* "bristle". However, it has been observed that the effect of lengthening is uneven among non-final vowels, in that penultimate vowels are generally longer [2, 3, 6].

One of the hypotheses invoked to account for the longer duration of penultimate vowels relative to antepenultimate ones is an overall compression effect due to number of syllables in the word [6]. This view is an example of broader "constituency effect hypothesis" according to which speakers tend to keep durations of larger phonological entities (e.g., phrases, words, feet, syllables) relatively constant; this implies that as a given entity contains more constituents (e.g., syllables in words and feet; words in phrases) the durations of these individual constituents are reduced [13]. Examples of constituency effect hypotheses are Lehiste's work on isochrony [5] and Campbell's on syllable duration [1].

The hypothesis in [6] correctly predicts that the [e] of *sétola* would be shorter than that of *séta*, since in the first case the word in which the vowel occurs has an extra syllable. However, this overall compression effect predicts also that the duration of [e] in *peséta* "peseta" would be the same as that of *sétola*, because they have the same numbers of syllables. The hypothesis cannot explain effects of the location of a syllable inside words of the same length. However, in a previous study it was shown that a vowel in penultimate position, such as the [e] of *peséta*, is not shorter than, say, the same vowel in *séta*. An alternative account based on foot structure and constraint interaction was therefore presented in [2].

We test here the effects of overall word length, position relative to the right edge of a word, and position relative to the left edge of a word in an experiment where we measure durations of stressed vowels in open syllables in all positions of two- and three-syllable words.

In addition to assessing these effects, we address an important methodological issue concerning the measurement of stressed word-final vowels. The duration of such vowels depends critically on the segmentation criteria selected by the labeler. This is due to the fact that stressed vowels in absolute final position are often "checked" in Italian, that is, they are followed by a glottal stop [15]. The impact of different measurement criteria is therefore evaluated and discussed.

2. METHODS

A database of 1062 stressed vowel tokens was the basis of our analyses, produced by a single speaker of Tuscan Italian. These vowels were the sole targets in (1062) carrier phrases containing nonsense target words, as in: *Cercami la parola pegn***á***pe sul dizionario* "Find the word *pegn***á***pe* on the dictionary" and *E' proprio p***á***pepa che volevo fare ieri sera* "It's really *p***á***pepa* that I wanted to do last night". Here, the bold-faced /a/ is the target vowel.

2.1. Statistical analysis

As in almost any corpus in which contexts and target identities vary on several factors, some degree of factor confounding is inevitable. That is, not all combinations of vowel identity, intra-word location, and postvocalic consonants identity occur equally often. In order to overcome this problem, we calculated in addition to the standard duration means also the *corrected means* [14]. The latter are computed as follows. Suppose that the factors on which the measured durations depend include segment identity (*ID*), intra-word location (*IWL*), intra-phrasal location (*IPL*), and stress (*STS*). The analysis method assumes that duration is given by:

$$\mathrm{DUR}(ID, IWL, IPL, STS) = \mathrm{CM}(IWL) \times \mathrm{B}(ID, IPL, STS). \quad (1)$$

The parameters *CM()* and *B()* can be estimated using standard Least squares techniques. By ensuring that the (geometric) mean of the *B()* parameters is 1.0, the parameters *CM()* can be taken as *estimates of the simple marginal means that we would have obtained in a perfectly balanced experiment*. The degree of difference between corrected and uncorrected means indicates the amount of

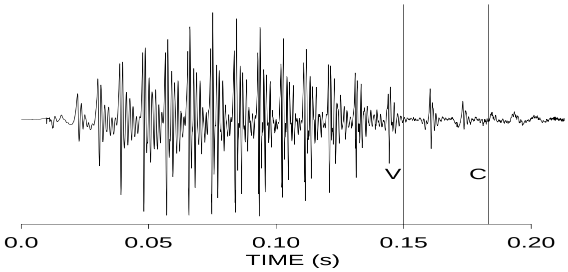

Figure 1: Waveform for a word-final stressed [a]. V = Offset of core vowel. C = offset of glottalization.

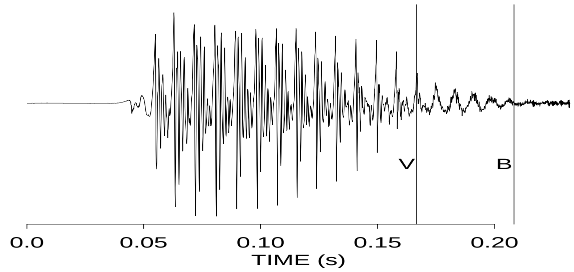

Figure 2: Waveform for a word-final stressed [a]. V = Offset of core vowel. B = offset of breathy voice region.

imbalance in the data, but can also be explained by a failure of the multiplicative model (1); if, on the other hand, the difference is small, than this strongly indicates that the data are relatively well-balanced and that the technique worked as intended. As it turned out, the latter was the case. We therefore only report results of the simple, uncorrected means.

All contrasts reported in this paper are statistically significant, due to the fact that with at least 300 observations in each condition the standard errors are always less than 2.5 ms. Even using conservative post-hoc contrasts (e.g., Scheffé intervals), 10 ms contrasts are always significant at $p < 0.01$.

2.2. Glottalization and finally stressed vowels

In Tuscan Italian, it appears that a glottal stop is often used to mark the end of a word-final stressed vowel when it occurs in an open syllable. Specifically, it has been proposed that the presence of a glottal stop serves the function of providing a coda consonant to an otherwise open syllable containing a short vocalic nucleus. By doing so, the weight requirement for stressed syllables, which need to be "heavy" [8], would be met by simply adding a coda consonant (the glottal stop) to the light vocalic nucleus [15].

Glottalization (also known as "laryngealization" or "creaky voice") is often found in the surroundings of a glottal stop. By glottalization we refer to a long and often irregular series of period-to-period intervals cause by forceful and sudden adduction of the vocal folds. Since it is often the case that stop closure is not complete, it is not rare that the only cue for the presence of a glottal stop consists of the glottalized region belonging to the vowel preceding it. In our data, glottalization was in fact found within the final portion of finally stressed vowels in open syllables (see Figure 1), as in the data presented in [15]. An alternative strategy seems to be adopted when the vowel is also in absolute phrase-final position [4]. In such a case, the vowel itself will result to be lengthened in Italian.

Irregular periodicity as a result of glottalization is also found in languages such as English and German, though it has a different purpose, i.e. that of marking the left edge of a vowel-initial word. This phenomenon seems to be correlated with higher level prosodic hierarchy effects. In [10] it is proposed that glottalization is favored at intonation phrase boundaries, even when no pause is present, as well as at main stress location within the target. In our data, it is possible that glottal stop presence is employed as a cue to mark the strong prosodic boundary that immediately follows a nuclear stressed word, but this is only very speculative at this point. In fact, the corpus target words were always (unless different instructions were given) pronounced with accentual prominence on them.

Apart from an abrupt change in periodicity, various acoustic marks of glottalization have been reported in the literature. The sudden closure of the glottis can produce increasing energy in the highest frequencies of the spectrum (especially in the second and third formant regions), although this effect is not always found. Another cue to glottalization is a relatively undamped waveform. In our data, however, word-final vowel offsets often presented acoustic indices that are not typical of glottalization, as described in the literature. In the same contexts where clear glottalization was found, a rather different phenomenon was often observed, i.e. a marked decrease in the amplitude of the pulses, which were only slightly irregular, and a more or less evident breathy quality. In this cases, a mild noise component in the higher frequencies was observed, which caused individual pulses to be hardly distinguishable. An example is shown in Figure 2.

As a consequence of the various marks of glottalization, a wide typology of word-final stressed vowels is included in our corpus. Among the various manifestations of these vowels, we find presence of period irregularity at offset or breathiness, which could optionally be followed by:

- A short period of silence followed by a low-frequency burst (around 2000 Hz). Note that the low frequency burst that was found in some of the cases appears to be due to the presence of the allophonic glottal stop.
- A short period silence, with no visible release burst.
- High-frequency frication noise, corresponding to the initial /s/ of the following word in the carrier sentence, and no visible glottalization.

In difficult cases, the power spectrum of the critical region was inspected, since a marker for creaky phonation consists of a relatively flat spectrum. When visible, the label was placed at the onset of the region presenting higher energy in the second or third formant. Breathy vocalic regions where recognized by the presence of high frequency random noise and by a drastic decrease of energy in the second formant region. The power spectrum of the breathy region was also often inspected, since a strong characteristic of breathiness is increasing spectral tilt. The label for the end of the glottalized/breathy region could immediately precede either the onset of silence (if a pause was produced between the target word and the next word in the carrier sentence) or the onset of high frequency frication noise (around 4000 Hz) associated with the first segment ([s]) of the following word in the carrier sentence.

Segmentation of the critical regions was conducted in a quasi-objective way, since not all cues for glottalization and/or breathiness are always present. What is more, increase in the period-to-period interval can also be related to other causes, such as an intonationally related, phrase-final pitch decrease. Given the relative difficulty in segmenting glottalized regions, clear instances were singled out by manual inspection of both waveform and spectrogram and were subsequently labeled by means of Entropic's Waves+.

In our analysis, the label corresponding to “core” vowel offset (V) was placed at the onset of the glottalized/breathy region, which is defined as the portion of the vowel where marked irregularity in the pitch periods or a drastic drop in fundamental frequency and/or energy was observed. We shall therefore present three alternative durations for word-final stressed vowels, that is the core vowel by itself (V), the core vowel plus the glottalized region (VC) and the core vowel plus the breathy region (VB).

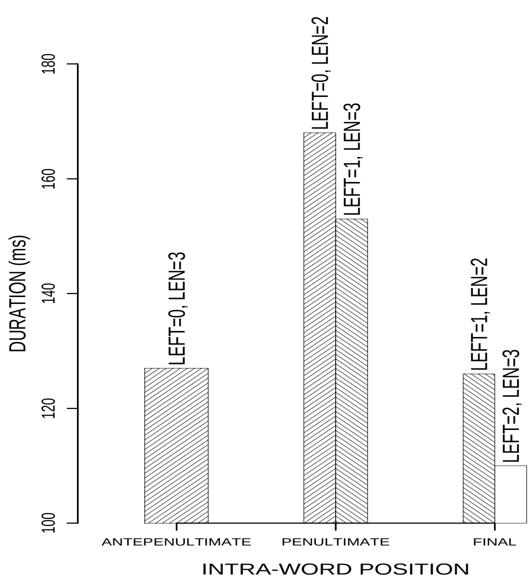

Figure 3: Mean vowel duration as a function of intra-word position, number of syllables to the left (LEFT) and word length (LEN).

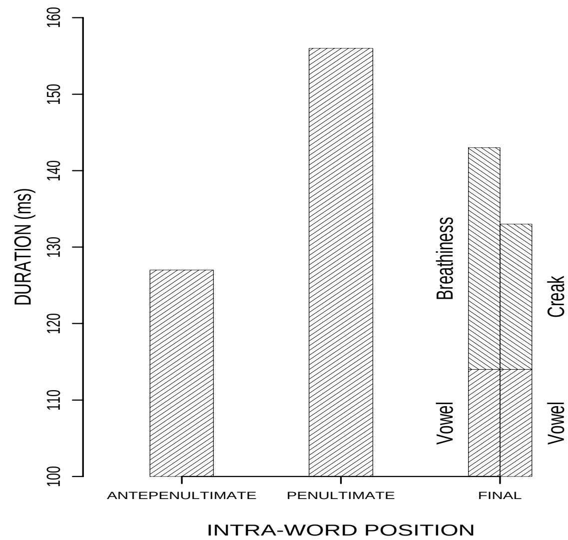

Figure 4: Mean vowel duration as a function of intra-word position. For word-final vowels, durations is divided into a pure vowel part and a creaky or breathy part.

3. RESULTS

Figure 3 shows that vowels in penultimate position are longer than vowels in antepenultimate or final positions. However, the picture is somewhat more complex when we take into account the number of syllables to the left of the target vowel and overall word length. We see that penultimate vowels are longer when they are also word-initial than when they are preceded by at least one syllable (168 ms vs. 153 ms). This effect can be describes as a *word length effect*, or, as in [12], as a *word-initiality effect*. The former may be more plausible, based on the fact that that vowels in final position are longer when they are preceded by 1 vs. 2 or more syllables; obviously, in neither position are these final vowels word-initial. Nevertheless, when we contrast word-initial antepenultimate vowels with non-word-initial penultimate vowels, there is a strong difference of 26 ms, even though these vowels both occur in words of length 3. In summary, these data can be best describes as involving two factors: word length and position with respect to the end of the word.

We now turn to Figure 4, where we plotted vowels in word-final positions in terms of the pure vowel, breathy, and creaky parts. The bars for antepenultimate and penultimate data are the same as the corresponding bars in Figure 3, except that for penultimate vowels we averaged the two cases kept distinct in that figure. The acoustic typology of word-final stressed vowels encountered in our database is of some independent interest. Among the 606 finally stressed vowels of the corpus, 366 presented a clear form of glottalization. Concerning these final vowels, we see that when we confine ourselves to the pure vowel part, word-final vowels are shorter (114 ms) than vowels in any other word position. However, when we include the creaky or breathy parts, their overall durations (133, 143 ms) – while still shorter than penultimate vowels (156) – become approximately slightly longer than vowels in antepenultimate positions (127 ms).

In summary, the data unambiguously shows that stressed vowels in open syllables are particularly long in penultimate position. This effect cannot be attributed to either word length effects or to a hypothetical word-initiality effect. However, there is an additional effect that causes durations of vowels in penultimate and word-final positions to shorten as these vowels are preceded by more syllables. This effect can be most parsimoniously described as a word length effect.

We note that these effects are sharply different from what is observed in American English [12], where penultimate stressed vowels in three-syllable words tend to be shorter than in either word-initial or word-final position. Similar results have been reported for German [7], and Mandarin Chinese [11]. Moreover, in both languages vowels in word-final syllables (that are not followed by any type of phrase boundary) tend to be longest of all.

4. DISCUSSION

In this paper we have shown that the duration Italian stressed vowels in open syllables depend on intricate ways on position in the word. We found that syllables in penultimate position are longer than syllables in other positions in same-length words, and that there is an additional effects which can be best described as an overall word length effect. We also showed that the duration of word-final vowels is not well-defined, because it critically depends on whether or not one includes in the vowel the final creaky or breathy region. If one does include these regions, then final vowels are not shorter than antepenultimate vowels. We note that glottalization in the final region of such vowels is very common, and the labeler has to make a choice whether or not to include the glottalized region within the vowel boundaries.

Though it has been previously argued that the glottalized region should be viewed as a syllabic coda that has the function to render the stressed syllable heavy [15], the facts are not as clear for the other manifestations of the final vocalic portion. Acoustic characteristics typical of breathy voice were found in the same location and in many instances. Such an effect had not been previously presented in the literature. Whether the breathy portion should be included or not in the vowel duration measurements is debatable. As we showed above, when the breathy portion of the vowel was included in the duration measurement, the word-final stressed vowel resulted to be longer than antepenultimate stressed vowels in open syllable. This is a problem for theories according to which all non-final vowels should be equally long.

Another interesting outcome of this study has to do with the potential effect of overall compression due to the size (in terms of number of syllables) of the entire word in which the vowel occurs. It had been previously argued that a possible explanation for the smaller duration of antepenultimate vowels was the effect of overall word compression [6], when stress initial words are considered. However, our data clearly show that such an effect, though its existence is suggested by the results, cannot explain the difference in duration between penultimate and antepenultimate vowels. This fact is especially clear when the duration of antepenultimate vowels (for which we only have data for stress initial words) and penultimate vowels in three syllable words are compared. Penultimate vowels are still much longer than antepenultimate ones (29 ms). Therefore, the origin of the positional effect must be of a different nature than a linear compression effect. A phonological account is not attempted here, while it is offered in [2].

5. CONCLUSION

Altogether, our results on stressed vowel duration in Italian challenge any model that invokes the use of a single rule of lengthening under stress in non-final open syllables. It was shown that antepenultimate vowels are always shorter than penultimate, even when penultimate vowels are non-initial in the word or when both are measured in three-syllable words. Also, the claimed short durations of word-final vowels depends critically on whether the final glottalized/breathy region is included in the measurements.

The relatively long duration of penultimate vowels reveals a timing pattern that is in a sense the opposite of what has been found in a diverse list of languages, including Mandarin Chinese, German, and American English. Here, the general rule seems to be that vowels are shortest in word-interior syllables, longer in word-initial syllables, and longest in word-final vowels. Clearly, a more cogent phonological account for these different timing patterns is needed.

REFERENCES

[1] Campbell, W., and Isard, S. 1991. Segment durations in a syllabic frame. *Journal of Phonetics 19* 37–47.

[2] D'Imperio, M., and Rosenthall, S. 1999. Phonetics and Phonology of Main Stress in Italian. *Phonology 16, 1*.

[3] Farnetani, E., and Kori, S. 1990. Rhythmic Structure in Italian Noun Phrases. *Phonetica 47*, 50–65.

[4] Helsloot, K. 1995. *Metrical Prosody*. Leiden: HIL Series.

[5] Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics 5*, 253–263.

[6] Marotta, G. 1985. *Modelli e misure ritmiche*. Bologna: Zanichelli.

[7] Möbius, B., and van Santen, J. 1996. Modeling segmental duration in German text-to-speech synthesis . In *Proceedings ICSLP* (Philadelphia, U.S.), pp. 2395–2399.

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

[9] Nespor, M., and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

[10] Pierrehumbert, J., and Talkin, D. 1992. Lenition of /h/ and glottal stop.s. In *Papers in Laboratory Phonology II*, H. Docherty and D. R . Ladd, Eds. Cambridge: Cambridge University Press, pp. 90–116.

[11] Shih, C., and Ao, B. 1996. Duration Study for the Bell Laboratories Mandarin Text-to-Speech System. In *Progress in Speech Synthesis*, J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, Eds. Springer-Verlag, New York.

[12] van Santen, J. 1992. Contextual effects on vowel durations. *Speech Communication 11* 513–546.

[13] van Santen, J. 1997. Prosodic modeling in text-to-speech synthesis. In *Proceedings of Eurospeech-97* (Rhodes).

[14] van Santen, J. Timing. 1997. In *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, R. Sproat, Ed. Kluwer, Boston, MA, ch. 5, pp. 115–139.

[15] Vayra, M. 1992. Phonetic explanations in phonology: laryngealization as the case of glottal stops in italian word-final stressed syllables. In *International Phonology Meeting* (Krems, Austria), pp. 275–294.