



The Duration Compensation Issue Revisited

Plínio A. Barbosa

Speech Prosody Studies Group, Dep. of Linguistics, State Univ. of Campinas, Brazil

pabarbosa.unicampbr@gmail.com

Abstract

This work presents a reanalysis of the duration compensation issue at the VC level, for which vowel duration is lengthened when replacing a following (long) voiceless consonant with a (short) voiced consonant. Analyses of fifteen Brazilian Portuguese words and pseudo-words reveal that (1) duration compensation is not restricted to consonants differing in voicing only, (2) duration compensation seems to be restricted to narrow-focussed words, and (3) duration compensation applies to the first following consonant only in VCC sequences. These results suggest that duration compensation could be a mechanism to learn how to plan the sequence of vowel onset positions at early stages of language acquisition.

Index Terms: duration compensation; speech timing; emphasis.

1. Introduction

Early studies showed that, at least in languages such as English, Swedish, and French [1, 2], the paradigmatic change of a voiceless obstruent by a homorganic voiced one causes a significant duration increase of the vowel preceding the voiced consonant. This vowel duration increase is later interpreted as compensatory at the VC level (see [3]) because the voiced obstruent following the lengthened vowel is shorter than its voiceless counterpart in those languages. In these same languages no duration compensation at the CV level is found, that is, there is no significant change in duration in the following vowel after a paradigmatic change between homorganic obstruents differing in voicing in the syllable onset.

Simple physiological explanations for the lengthening of the vowel following a voiced obstruent, such as to make easier the pronunciation of this obstruent, happened to be wrong because the effect is seemingly language-specific. Yet the reasons for this compensation are not completely understood. Experimental evidence allowing the interpretation of this phenomenon in terms of general rhythmic constraints applying onto the C-V transitions sequence is presented in this paper.

Languages often cited as not exhibiting duration compensation at the VC level include Arabic [4, 5], Polish, Czech [6], and Japanese [5], at least in part. As for Arabic, Mitleb's study with the Jordanian variety did not find evidence for duration distinction between the long open vowel /a:/ in two contrasting words, "Kaaz" vs "Kaas", where /s/ and /z/ do not contrast in duration as is the case for English, Swedish, and French. Port et al.'s study with the Egyptian, Iraqi and Kwatian Arabic varieties, on the other hand, presents a more complex picture. Even though both short and long /a/ preceding the voiced alveolar [d] are 13 ms significantly longer than those same vowels preceding the homorganic voiceless [t], no significant difference in duration was found for the contrasting consonants. A more recent work with Ammani Arabic [7] confirms this weak effect and re-

lates it to the absence of duration contrast for segments differing in voicing only. Another crucial and intriguing fact in [5] is that the phonological change of quantity (short vs long /a/) does not affect the following consonant duration. In Japanese, the duration change of both preceding and following consonants, when replacing them by their voiced counterparts, causes lengthening of the vowel [5]. Not all these studies are contradictory with the duration compensation issue at the VC level because (1) in at least one study the following consonants do not contrast in duration, as it seems to be the case in [4]; (2) the duration distinction is too small and would deserve another evaluation, as shown in [5] for Arabic; (3) Japanese presents compensatory duration, although it seems to be bidirectional.

Polish and Czech seem to be actual counter-examples. In fact, Keating [6] showed that, in both languages, although the homorganic alveolar plosives contrast in duration, preceding low, central vowels exhibit no duration distinction at all. The tested words were recorded in isolation, and seemingly uttered with emphasis. Although emphasis realisation certainly affects the timing of the segments in the word, possibly modifying compensatory effects, emphasis cannot explain the absence of compensatory effect at the VC level because, as it will be shown here, emphasis makes the phenomenon more apparent, at least in Brazilian Portuguese. In the following, a proposal for considering the compensatory effect as a consequence of more general rhythmic constraints operating in the utterance and possibly learned in the first stages of language acquisition will be presented.

2. A rhythmic hypothesis for the VC duration compensation issue

When comparing prosodically similar utterances in English, Swedish and French, the sequence of vowel onset positions in the being-compared utterances tends to be relatively fixed. Research on the perceptual-center phenomenon (cf [8, 9, 10], *inter alia*) can be evoked to suggest that a reason for that behaviour is the need to not disturb the perception of (the sameness of) speech rate across utterances [10]. This hypothesis will be called here the "rhythmic hypothesis". As a consequence of it, any paradigmatic change of a voiceless obstruent by its shorter, voiced counterpart should cause a complete compensatory effect on the preceding vowel, *ceteris paribus*. That this compensation needs to take place at the VC level is apparent, since no change in the vowel preceding the voiced obstruent would cause the next vowel in the utterance to occur earlier in time, which would change the perception of speech rate across similar utterances.

If the rhythmic hypothesis is correct, the aforementioned effect would take place for any contrasting consonants differing in duration, not only between single-feature contrasts such as voicing. That is, if the compensatory effect takes place due

to consonant-related durational differences, the only contrast needed is the one of intrinsic duration. Thus, any paradigmatic change from a long C segment to a shorter one, would cause a compensatory effect on a preceding vowel. As additional pieces of evidence in favour of the rhythmic hypothesis are the results of early work pointing out the relevance of the control of the realisation of the sequence of CV transitions in speech production and of the corresponding tracking of these same transitions in speech perception [11, 12]. This would imply that, when the speaker does not intend to signal any change of speech rate to the listener, s/he will maintain CV transitions temporally stable. This would explain the early findings on duration compensation.

Alternatively, the work by [13] shows, on one hand, that segment duration depends on several other factors such as lexical stress, prominence level, position in the sentence, among others. On the other hand, their findings did not show evidence for the syllable as a unit within which segment duration compensation would take place. The work presented here controls for the aforementioned variables that do affect segment duration by focussing on the syllable-sized framework. As for the second issue, what the authors seem to ignore is that the V(.)C unit, which is a phonetic syllable, is the appropriate unit for segment duration compensation, instead of the phonological syllable.

In the work presented here, the rhythmic hypothesis is assessed with the recordings of 15 words (and pseudo-words) from two subjects of different dialects of Brazilian Portuguese (henceforth BP). The main goal of this experimental setting is to explore a larger set of durational contrasts for the following consonants than the one explored earlier in the literature. The results presented here show that there is duration compensation at the VC level, that this compensation takes place with a paradigmatic change of any two consonants contrasting in duration, but that compensation seems to apply to narrow-focussed words only.

3. Methodology

Two kinds of possible duration-related causes for compensation in BP were analysed in this work: (1) vowel duration compensation due to the duration contrast of following consonants, and (2) duration compensation due to an increasing number of following consonants. Two corpora were used for investigating each one of these possible causes.

The first corpus is the reading, by a female speaker from São Paulo State, aged 40 at the time of recording, of eight words and pseudo-words (henceforth *target*) inserted in the carrier sentence *Não sei se a target mascara o efeito desejado* ((I) don't know whether the *target* masks the intended effect.) The sentences were read five times in random order by block in three prosodic conditions: broad focus in the whole utterance (N, neutral), narrow focus on the target (E1, emphasis 1), and enhanced narrow focus on the target (E2, emphasis 2). The broadly-focussed utterance was presented by using a capital letter only for the first character of the 12-point printed sentence. Both degrees of emphasis were elicited by using capital letters for all characters of the target words/pseudo-words, with 14-point printed characters for E1, and 20-point printed characters for E2. The speaker was simply informed of this difference and invited to reproduce it orally. Pseudo-words were read as naturally as words.

All targets (pseudo-words signalled by *) have penultimate stress, which is the default lexical stress pattern in Por-

tuguese [14] (translation, medial VC unit): *casa* (house, /az/), *caça* (hunt, /as/), **caba* (/ab/), *capa* (coat, /ap/), *casta* (caste, /ast/), *cata* ([s/he] picks, /at/), *caspa* (dandruff, /asp/), and *cara* (face, /ar/). Several possibilities of contrast are found in this corpus: between voiceless and voiced obstruents following the stressed vowel (*casa* vs *caça* and *caba* vs *capa*), between non-homorganic consonant intervals differing in duration (*capa* vs *cara* or *casa*). The vowel /a/ was chosen to occupy both V positions because it is the most frequent vowel in BP. The three prosodic conditions also differed with respect to fundamental frequency (F₀) excursion, with a high tone on the target with a range that is higher for E1, and still higher for E2 (both accompanied by lowered F₀ values post-focally).

The second corpus introduces a set of seven target words (and pseudo-words signalled by *) differing in number of intervocalic consonants (from /C/ to /NCCC/), evaluated under four distinct prosodic conditions. The targets are: *passa* ([s/he] passes), *pata* (paw), *pasta* (suitcase), **panta*, **patra*, **pastra*, **panstra*. Each target was read in the carrier phrase composed of two parts, as follows. The phrase *Pedro não sabe da* (Peter doesn't know about the) precedes the target, which is then followed by distinct linguistic materials, in order to elicit four prosodic conditions, as such: (1) "target VERMELHONA" (extremely red): deaccented condition (D); (2) "target *vermelhona*": neutral condition 1 (N1); (3) "target, *vermelha ou não*" (red or not): neutral condition 2 (N2); and (4) "TARGET *vermelhona*": emphatic condition (E). Conditions N1 and N2 contrast two degrees of prosodic strengthening, since there is a stronger boundary between the target and the following word for N2. The sentences were read in random order by block eight times by a female speaker from Mato Grosso do Sul State (Southwestern Brazil), aged 35. If any compensatory effect is expected, the stressed vowel will be the shorter, the higher the number of intervocalic consonants. The choice of female speakers was motivated by the fact that duration compensation was found very strong in a male speaker [15] and we would like to test the validity of the phenomenon in female speakers.

For both corpora, the durations of all segments of the targets and of the first consonant of the following word were measured and taken as dependent variables in completely randomised, 2-Way ANOVA tests. Segment intervals were delimited according to F2 onset and offset in broad-band spectrograms for vowels and nasal murmurs. Coda [s] offset was marked at the end of friction energy in high frequencies (above 4 kHz), whereas [t] offset in /tr/ syllables was defined by the F2 onset of the vocalic part that precedes the acoustic segment corresponding to the alveolar beat of the tap. For statistical analysis, two factors were chosen: factor CONDITION with three levels for the first corpus and four for the second one, and factor WORD, with eight levels for the first corpus, and seven for the second one. Scheffé post-hoc tests were carried out to evaluate differences between groups. In all cases the adopted significance level was 0.01. These ANOVA models were used to evaluate the general behaviour and the differences across different levels of the CONDITION factor. Corrected effect sizes were computed for each factor to evaluate the respective contributions for explaining the duration variance.

Linear regression models were fitted to quantitatively evaluate the degree of compensation in the conditions where it takes place. A complete compensatory effect would be represented by the equation $durC + k.durV = constant$ with $k = 1$. Fitted linear regressions are of the form $durC = -k.durV + constant$. Full compensatory effects are found for k approaching the unitary value. Partial compensatory ef-

fect would be found for $0 < k < 1$.

4. Results

As for the first corpus, which evaluates the relation between duration compensation and the nature of the following consonant, there was an interaction between the two factors only for the lexically stressed vowel (/V/) and the second onset consonant (/C₂/). The general trend was the splitting of mean duration, for all segments, into two separate groups, one for the N condition, and the other for E1 and E2 taken as one single group. The results of the ANOVA test for the first corpus for factor WORD can be seen in Table 1 for the stressed vowel and the immediately following consonant.

Table 1: Scheffé distinct groups for segment durations among levels of factor WORD (indicated by the VC(C) sequence) in the first corpus. The first row indicates the groups separated according to the duration of the stressed vowel, whereas the second row indicates the groups separated according to the duration of the first consonant in the VC(C) sequence.

segment	word groups
/V/	ap, at, ab, asp, as, ast < asp, as, ast, ar < az, ar
/C ₁ /	ar < ast, asp, ab, az < at, ap < as

As regards duration compensation in the three conditions pooled together as shown in Table 1, observe that (1) the longest stressed vowels are those followed by [z] (in the word *casa*) and [r] (in the word *cara*), which are the shortest consonants of the set; (2) the longest medial consonant is [s] (in the word *caça*), and the duration of its preceding vowel is in the two groups with the shortest durations; (3) stressed vowels and following voiceless plosives ([p],[t]) are placed respectively leftwards and rightwards in terms of duration; (4) there is clearly an effect of the number of segments in word-medial position, since the voiceless plosives are shorter when preceded by coda /s/ than when immediately preceded by the stressed vowel.

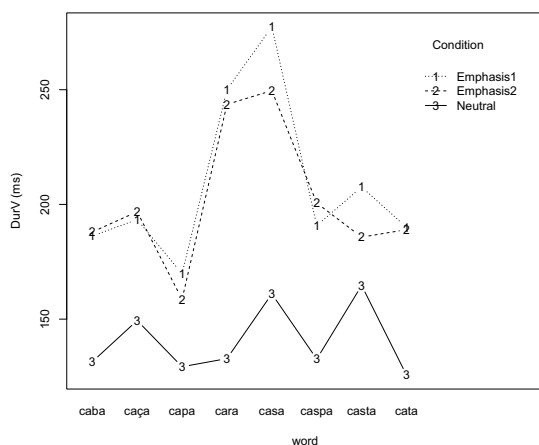


Figure 1: Stressed vowel duration mean for factor WORD in the first corpus. See text for details.

Figure 1 presents the stressed vowel mean duration for all

targets separated according to the factor CONDITION. Compare in this figure the mean duration positions in the y-axis with the order of mean duration consonant in Table 1. It can be observed that the emphatic conditions (either E1 or E2) are crucial to enhancing the differences (see the higher means of /a/ duration preceding [z] and [r]). This enhancing of the phenomenon for emphatic words can explain the previous findings in the literature, working mainly with isolated words.

The 2-Way ANOVA models for the stressed vowel and the following consonant explain respectively 64 % and 87 % of the duration variance for these two segments. Corrected size effects (ω^2) for the stressed vowel duration are 41 % for the CONDITION factor and 23 % for the WORD factor. As for the following consonant, ω^2 are 25 % for the CONDITION factor and 62 % for the WORD factor. With the VC(C) sequence duration as a dependent variable, the CONDITION factor explains 50 % of the duration variance whereas the WORD factor explains 36 % of the duration variance.

Considering the three prosodic conditions, the only regression model explaining more than 20 % of the variance of the data was the model between the vowel duration and the immediately following consonant for the two emphatic conditions pooled together (the neutral condition, even after excluding three outliers, produced a model explaining lesser than 19 % of the consonant duration variance with a non significant slope coefficient). The model for the emphatic targets is ruled by equation: $durC = -0.92durV + 331$ (ms), with $R^2 = 35\%$ and p at least $< 10^{-8}$ for both coefficients. The model taking the whole duration of the C intervocalic interval had a determination coefficient $R^2 < 6\%$. The relationship between the duration of the vowel /a/ and the following consonant can be seen in Fig.2.

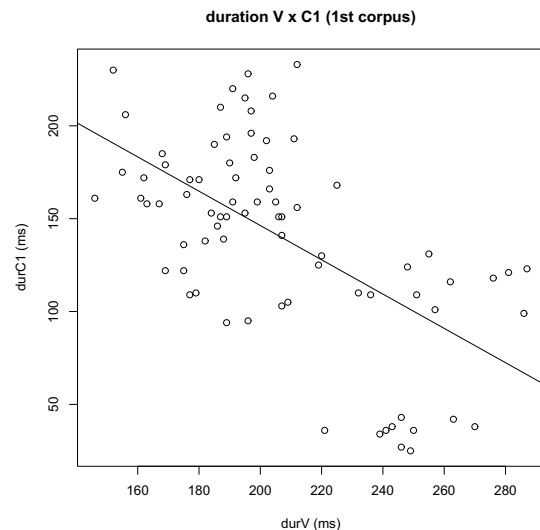


Figure 2: Scatter plot showing vowel duration (x-axis) and consonant duration (y-axis) in the first corpus pooling the two emphatic conditions together.

In comparison with the first corpus, the second one keeps only one degree of emphasis, while adding a prosodic boundary strength contrast (N1 vs N2), a deaccented condition (D), and the inclusion of the nasalised vowel [ẽ]. No interaction between the factors was found for any measured segment. As for factor

CONDITION, three or four levels are usually distinguished.

Table 2: Scheffé distinct groups for segment durations in the VC₁(C) sequence among levels of for factor WORD in the second corpus.

segment	word groups
[a]	at, atr < ast, as, astr
[ē]	ant < anstr
/C ₁ /	anstr, ast, ant, astr < atr, at < as
VC ₁ (C)	at, ant < atr < as, ast < astr, anstr

As for the results for factor WORD, shown in Table 2, observe that (1) the number of segments affect the overall VC₁(C) duration: in general, the greater the number, the longer the VC(C) unit (row 4). Nevertheless, this result is not common to all targets, since the VC₁(C) units in the targets *pata* and *panta* are grouped together, despite the intrinsically longer duration of the nasalised vowel; (2) the voiceless plosive [t] (row 3) is shorter when preceded by the coda [s] or the archiphoneme /N/; (3) the duration of coda [s] does not change, despite the increasing number of segments in its VC₁(C) unit (row 3).

The 2-Way ANOVA models for the stressed vowel and the following consonant explain respectively 87 % and 84 % of the duration variance for these two segments. Corrected size effects (ω^2) for the stressed vowel duration are 67 % for the CONDITION factor and 20 % for the WORD factor. As for the following consonant, ω^2 are 40 % for the CONDITION factor and 44 % for the WORD factor. With the VC(C) sequence duration as a dependent variable, the CONDITION factor explains 61 % of the duration variance whereas the WORD factor explains 30 % of the duration variance.

Among the four prosodic condition in the second corpus, only the emphatic one produces a linear regression model with a determination coefficients higher than 20 % (when pooling the other three conditions together or taking them separately, the best models for them produced determination coefficients lesser than 5 %). After excluding two outliers for ensuring normality of residuals and homecedasticity between residuals and predict values, the model taking the vowel duration and the following consonant was ruled by equation: $durC = -0.59durV + 308$ (ms), with $R^2 = 36$ % and p at least $< 10^{-5}$. The model taking the whole duration of the C intervocalic interval had a determination coefficient $R^2 < 6$ %. The relationship between the duration of the vowel /a/ and the immediately following consonant can be seen in Fig.3.

5. Discussion

An almost full compensatory effect ($k = 0.92$) was found between the duration of the stressed vowel and that of the immediately following segment for targets read emphatically in the first corpus. Emphasis in a word is the situation most closely related to the pronunciation of isolated words. Considering that only consonant C₁ has its duration compensated in the preceding vowel in the emphatic condition, but not the whole consonantal interval, the rhythmic hypothesis seems not to hold for CVCCV words. The absence of compensation in this word structure would cause vowel onsets not be aligned in prosodically similar utterances differing in duration in the CC sequence, which is contrary to the rhythmic hypothesis. The linear model suggests that compensation needs only a contrast in duration between paradigmatically comparing C₁ consonants. A striking

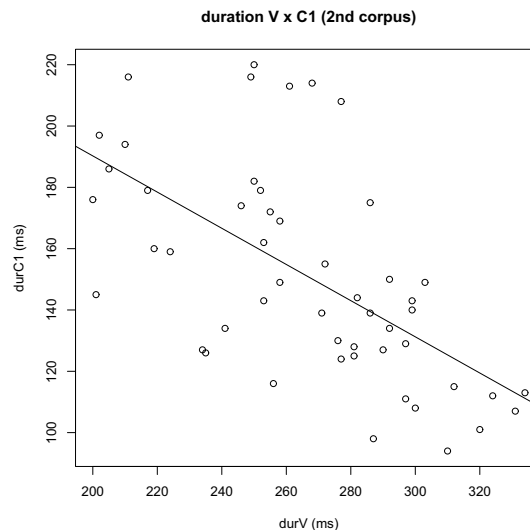


Figure 3: Scatter plot showing vowel duration (x-axis) and consonant duration (y-axis) in the second corpus for the emphatic condition.

example of this fact, is the contrast between the shortest consonant ([r]) and any other consonant in emphasised words, which produces the reversed effect on the preceding vowel (the vowel [a] preceding [r] is the longest in emphasised words).

A partial compensatory effect was found between the duration of the stressed vowel and that of the immediately following segment in the second corpus, for which the main changes were prosodic condition and number of intervocalic consonants. Also for this second corpus, targets read emphatically were the only ones exhibiting the aforementioned partial compensatory effect ($k = 0.59$). Although the rhythmic hypothesis seems not to hold between the vowel and the following consonant interval as a whole, a weak compensatory effect is found between the voiceless plosive [t] and the preceding vowel when [t] is replaced by the consonant intervals [st] and /Nt/.

Taking into account the findings shown here, the compensation seems to be partial and restricted to emphatic words, although stronger for CVCV words. Considering these canonical word structures in BP, preservation of VC-level duration is in principle compatible with a hierarchical, perturbational analysis of timing where syllable-sized durations behave as a stable frame within which subsyllabic segments are fit [16, 17, 18], at least considering CV syllable sequences in early acquisition. This is also compatible with these theoretical accounts because emphatic words and hyperarticulated words are the usual situation in language acquisition, during which the simplest word structures are the first to be produced [19]. Compensation at this level of learning suggests preservation of vowel onset positions along the utterance, which probably contributes to the acquisition of rhythm.

6. Acknowledgement

We acknowledge a grant number 301387/2011-7 from CNPq, and thank the three anonymous reviewers for the thoughtful suggestions.

7. References

- [1] Chen, M., "Vowel length variation as a function of the voicing of the consonant environment", *Phonetica*, 22:129–159, 1970.
- [2] Lehiste, I., *Suprasegmentals*, Cambridge, Massachusetts: The MIT Press, 1970.
- [3] Maddieson, I., "Phonetic Universals", in W. J. Hardcastle and J. Laver [Eds], *The Handbook of Phonetic Sciences*, 619–639, Oxford: Blackwell, 1997.
- [4] Mitleb, F. M., "Voicing effect on vowel duration is not an absolute universal", *J. Phon.*, 12:23–27, 1984.
- [5] Port, R., Al-Ani, S. and Maeda, S., "Temporal compensation and universal phonetics", *Phonetica*, 37:235–252, 1980.
- [6] Keating, P., "Universal Phonetics and the Organization of Grammars", in V. A. Fromkin [ed], *Phonetic Linguistics. Essays in honor of Peter Ladefoged*, 115-132, Orlando: Academic Press, 1985.
- [7] Zawaydeh, B. A. and de Jong, K., "Stress, phonological focus, quantity, and voicing effects on vowel duration in Ammani Arabic", in *Proc. 14th ICPhS San Francisco*, 451-454, 1999.
- [8] Marcus, S. M., "Acoustic determinants of Perceptual center (P-center) location", *Perception & Psychophysics*, 30 (3), 247-256, 1981.
- [9] Morton, J., Marcus, S. M., Frankish, C., "Perceptual centers (P-centers)", *Psychological Review*, 83 (5), 405-408, 1976.
- [10] Pompino-Marschall, B., "The syllable as a prosodic unit and the so-called P-centre effect", *Forschungsberichte Institut für Phonetik und Sprachliche Kommunikation der Universität München*, 29:66-124, 1991.
- [11] Dogil, G. and Braun, G., *The PIVOT model of speech parsing*, Wien: Verlag, 1988.
- [12] Pompino-Marschall, B., "On the psychoacoustic nature of the P-Center phenomenon", *J. Phon.*, 17:175–192, 1989.
- [13] van Santen, J., Shih, C., "Suprasegmental and Segmental Timing Models in Mandarin Chinese and American English", *Journal of the Acoustical Society of America*, 107 (2):1012-1026, 2000.
- [14] Cintra, G., "Distribuição de padrões acentuais no vocábulo em português", *Confluência*, 5 (3):82–93, 1997.
- [15] Barbosa, P. A. "At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration", in *Proc. First ESCA TRW on Speech Production Modeling: From Control Strategies to Acoustics; 4th Speech Production Seminar: Models and Data* (pp. 85-88). Autrans, France, 1996.
- [16] MacNeilage, P. F., "The frame/content theory of evolution of speech production", *Behavioral and Brain Sciences*, 21:499-511, 1998.
- [17] Cummins, F., Port, R., "Rhythmic constraints on stress timing in English", *J. Phon.*, 26:145–171, 1998.
- [18] Fischer-Jørgensen, E., "Sound duration and place of articulation", *Zeits. für Sprachwissenschaft und Kommunikationsforschung*, 17:175–207, 1964.
- [19] Kent, R., Miolo, G., "Phonetic abilities in the first year of life", in P. Fletcher and B. MacWhinney [Eds], *The Handbook of Child Language*, 303-334, Oxford: Blackwell, 1995.