ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Intrinsic Prosodic Properties of Stressed Vowels in European Portuguese

Francisco Costa

ILTEC – Instituto de Linguística Teórica e Computacional
Lisbon, Portugal
fcosta@iltec.pt

Abstract

The present paper describes a study developed with the purpose of inspecting the intrinsic acoustic properties of vowels of stressed syllables in European Portuguese (with minor considerations about semi-vowels as well). The features we studied are fundamental frequency and duration. As in other languages, to some extent, these properties depend on vowel quality and context. The study addresses this issue. It is shown that not all correlations that have been discovered for other languages hold as well, but nevertheless these properties do differentiate vowels and diphthongs, and, interestingly, in some cases the lack of those correlations is offset by other correspondences.

1. Introduction

It has been stated for a number of languages that, all other factors being equal (segmental and prosodic context), a high vowel will display values for certain prosodic characteristics that will be somewhat different from the ones associated to a low vowel [9]. Fundamental frequency and duration have been identified as being among the relevant prosodic features. In this sense, we can speak of intrinsic, or inherent, fundamental frequency (henceforth IF0) and duration. Since this difference recurs in a number of languages, it has also been hypothesized that this pattern is to some extent universal [14].

The importance of this kind of micromelodic variation is revealed, for instance, when studying prosody at larger levels, because we need to know whether the differences we observe are due to prosodic context or inherent features. Because of this, the study of intrinsic prosodic properties can be an important first step to prosodic analyses, especially when the language we want to describe is not fully documented as far as segmental variation in inherent properties is concerned.

This is the case of European Portuguese (hereafter E.P.). Some studies do exist, but they do not focus on IF0 aspects [1-3,6,13]. Brazilian Portuguese has already been investigated [12]; however, it is legitimate to suspect that there may be differences between the two varieties, for they are quite distinct in several phonetic aspects.

For a number of languages, it has been noticed that F0 and duration vary with vowel height (high vowels are consistently shorter than low ones and display higher F0) [14], and in some languages they also seem to be constrained by place of articulation [12]. Also, some contextual aspects seem to affect these values: F0 is higher on vowels after a voiceless obstruent than on vowels following a voiced consonant [9], and vowel length is affected by the voiced/voiceless nature of the postvocalic consonant (longer vowels before voiced consonants) [13], among other factors.

In this study we concentrate on stressed vowels and diphthongs only. The behavior of unstressed syllables is not so important as far as IF0 and inherent duration are concerned (when one has macrolevel prosodic considerations in mind), because the nucleus of larger prosodic domains will only coincide with stressed syllables.

Besides intrinsic F0 and duration, the effect of certain contextual factors on these variables is also examined, namely segmental context (neighboring consonants only) and syllable structure (rhyme only, onset was assumed to be irrelevant).

2. Data and Method

We created a *corpus* specifically for this task. The *corpus* is composed of 228 words occurring in a sentence of the type «Eu digo *x* sempre» (lit. «I say *x* always»), where *x* is the word in which the vowels to be studied are included («Eu digo *x* agora» - «I'm saying *x* now» - would be more natural, but because the word following *x* begins with a vowel, *sandhi* would occur between this vowel and the last vowel of *x*, which might influence results). The fact that these words all occur in the same prosodic context (the one dictated by the template sentence above) guarantees that differences in the prosodic features at stake are not due to different prosodic contexts. This also explains the position of the word in the sentence: it assures that these words were uttered in a prosodically neuter context (they were not the nucleus of an intonational phrase).

We built the word list in such a way as to allow the comparison of the factors (at the segmental and syllabic level) that we suspected might influence fundamental frequency (hereafter F0) and duration, inspired by general trends that have been revealed for other languages (for instance, it is widely mentioned that F0 is higher after voiceless obstruents than after voiced ones [e.g., 9]; therefore, specific pairs of words were included to test this).

Since most Portuguese words are stressed on the penultimate syllable, most words in our *corpus* present this stress pattern, too, in order to reduce variability (so that results can be compared). The ones that do not obey this restriction were included in the *corpus* in order to test the effect of stress or because they present characteristics that we want to test but that cannot be found on paroxytones. However, the first group turned out to be too small to draw conclusions from, and so the second group was not included in this presentation either. Therefore, all words considered are paroxytones. Most of the *corpus* elements also have CV stressed syllables (since this is the unmarked case), except where noted. The final vowel was also generally constant ([ɐ], since all others can undergo deletion in some situation or other), assuring that differences in measured values are not due to differences in the following vowel.

Only when no word could be found with [ɐ] was a word with another vowel used (this was exceedingly rare): care was taken so that all test words were existing, 'common' words (since we were already using read speech, we did not want to include another factor that might increase artificiality). Word length did vary asystematically however, but nevertheless most test elements were two syllable words.

We asked six subjects (three female, three male) to read all 228 sentences aloud and recorded them. They were all around the same age (early twenties). They were native speakers of the same regional variety (Standard E.P.), and so were their parents.

A high quality microphone was employed and the distance from it to each subject's mouth was constant and equal for all subjects. The recordings were performed inside an anechoic chamber, so as to obtain as little sound interference as possible. The sound was digitized at a frequency range of 11025Hz (sampling rate of 22050Hz, 16 bits).

Each participant was asked to read the said set of sentences. These were displayed on a laptop computer screen at random order and in sequence. We included five additional sentences (following the same template) at the beginning of the slideshow. These were subsequently discarded. This precaution was to avoid the list effect. The last sentences were kept, since the participants had no way of knowing how many sentences were left at any given moment.

After the data were collected, all the words obtained were analyzed: the duration of the nucleus of each word's stressed syllable was measured, as well as the fundamental frequency value aligning with the center of its vowel's stable part (E.P. is not a tonal language, so we need not care about F0 movements in a single vowel).

This acoustic analysis was carried out using Sensimetrics SpeechStation2. F0 was measured in Hz and duration in milliseconds. Unfortunately, some F0 values were not correctly identified by the F0 detector (which sometimes reported absurd readings - other programs were tested, but they also performed poorly). Since this problem seemed to depend on subject specific characteristics (F0), we had no choice but to exclude these words from further consideration about fundamental frequency.

3. Results

We divided the words in our *corpus* in several groups, according to the stressed syllable's structure (the presence or absence of coda, nucleus structure) and the nasal or oral nature of its nucleus.

In Table 1 we present the results for words with a stressed syllable of type CV (with a non-nasalized vowel); Table 2 relates to CVC stressed syllables (again with an oral vowel); Table 3 compares the mean vowel length of vowels in stressed CV to the one of vowels in CVC structures (oral only); Table 4 displays the behavior of nasal vowels (in CV stressed syllables) – note that there are no low nasal vowels in Standard E.P –; Table 5 depicts the effect of the neighboring consonants on the stressed vowel's length and F0 (only the vowel [ɛ] was considered and it is assumed that the other vowels behave in a similar way; also, point of articulation was not investigated, but delegated to future work – some studies do point out that velar consonants influence neighboring vowels in ways that other consonants do not [13]). We also investigated diphthongs' length, but because it is difficult to discriminate the boundary between a glide and a vowel in a spectrogram, we decided to include pairs of words in our *corpus* such that they only differed in that one of them contained an additional semivowel adjacent to the stressed vowel. This enabled us to measure glides' length by subtracting the vowel's duration from the diphthong's. Table 6 presents these results. In all these tables, boldface denotes statistical significance at the 0.05 level or better. When values are compared to the overall mean, this measurement pertains to z-tests; when values are relative to a certain characteristic (syllable type in Table 3, nasality in Table 4, and presence vs absence of semivowels in Table 6), it was obtained with both Student's t and Tukey-Kramer tests.

Table 1: *Mean F0 (in Hz) and duration (in ms) of oral vowels in a stressed CV syllable (*v*). $\bar{x}$ denotes the mean value across all vowels.*

Vowel	Duration		F0 (female)		F0 (male)	
	v	v/ $\bar{x}$	v	v/ $\bar{x}$	v	v/ $\bar{x}$
[i]	111	**0.87**	220	**1.05**	126	**1.04**
[e]	125	0.97	209	1.00	120	0.99
[ɛ]	139	**1.08**	200	**0.96**	119	0.98
[u]	110	**0.86**	226	**1.08**	126	**1.04**
[o]	129	1.01	210	1.00	120	0.99
[ɔ]	139	1.08	202	**0.97**	120	0.99
[ɐ]	132	1.03	203	0.97	119	0.98
[a]	141	**1.10**	200	**0.96**	118	0.98
$\bar{x}$	128	1	209	1	121	1

Table 2: *Mean duration (in ms) and F0 (in Hz) of oral vowels in CVC stressed syllables (represented as* v*) and mean of all vowels in this context (*$\bar{x}$*).*

Vowel	Duration		F0 (female)		F0 (male)	
	v	v / $\bar{x}$	v	v / $\bar{x}$	v	v / $\bar{x}$
[i]	95	**0.81**	225	**1.07**	128	1.05
[e]	132	**1.12**	206	0.98	125	1.02
[ɛ]	124	1.05	206	0.98	121	0.99
[u]	89	**0.75**	225	**1.07**	127	**1.04**
[o]	115	0.97	210	1.00	121	0.99
[ɔ]	135	**1.14**	202	**0.96**	121	0.99
[ɐ]	118	1.00	205	0.97	114	0.93
[a]	134	**1.14**	206	0.98	117	0.96
$\bar{x}$	118	1	211	1	122	1

Table 3: *Mean duration (in ms) of oral vowels according to syllable type. The right column represents the relative distance between these values.*

Vowel	CV syllable	CVC syllable	Ratio
[i]	111	95	**0.86**
[e]	125	132	1.06
[ɛ]	139	124	**0.89**
[u]	110	89	**0.81**
[o]	129	115	**0.89**
[ɔ]	139	135	0.97
[ɐ]	132	118	0.89
[a]	141	134	**0.95**
$\bar{x}$	128	118	**0.92**

Table 4: *Mean duration (in ms) of oral vowels vs. nasal vowels in a stressed CV syllable. The last row presents the overall trend.*

Vowel	Oral Vowel	Nasal Vowel	Ratio
[i]~[ĩ]	117	175	**1.50**
[e]~[ẽ]	132	193	**1.46**
[u]~[ũ]	118	171	**1.45**
[o]~[õ]	139	193	**1.39**
[ɐ]~[ɐ̃]	144	193	**1.34**
$\bar{x}$	130	185	**1.42**

Table 5: *The effect of neighboring consonants on the duration of vowel [E] in a stressed CV syllable.* v *denotes mean vowel length after/before the corresponding consonant.*

Consonant Class	Context			
	CV		V.C	
	v	v / $\bar{x}$	v	v / $\bar{x}$
voiceless plosive	137	0.99	120	**0.86**
voiced plosive	160	**1.15**	152	1.09
voiceless fricative	135	0.97	133	**0.96**
voiced fricative	148	**1.06**	162	**1.17**
lateral	146	**1.05**	145	1.04
nasal stop	144	1.04	142	1.02
flap	136	0.98	166	**1.19**
voiceless obstruent	136	0.98	126	**0.91**
voiced obstruent	153	**1.10**	157	**1.13**
sonorant	144	**1.04**	151	1.09
voiced consonant	148	**1.06**	154	**1.11**
plosive	146	1.05	134	**0.96**
fricative	141	1.01	146	1.05
obstruent	145	**1.04**	140	**1.01**
$\bar{x}$	139	1	148	1

Table 6: *Mean duration (in ms) of semivowels [j] and [w] in prevocalic and postvocalic position. Results are for oral stressed diphthongs occurring in an open syllable only.* GV *and* VG *denote the duration of a glide plus vowel sequence or a vowel plus glide sequence respectively.*

Context, Segment	V	GV	GV-V	GV/V
[j]V	132	200	68	**1.52**
[w]V	137	236	99	**1.72**
mean	134	217	83	**1.61**
Context, Segment	V	VG	VG-V	VG/V
V[j]	153	183	30	**1.19**
V[w]	182	213	30	**1.17**
mean	169	199	30	**1.18**

3.1. Analysis

Note that we have only presented F0 results in the first two tables. We have in fact computed these values for the other data as well, but found no interesting effect on F0 (it remains more or less constant independently of the existence of a syllable coda, the vowel's nasalized quality or even the manner of articulation of the adjacent consonants or the state of the glottis).

The fuzziness of these results is more or less expectable: we cannot assume too much precision in articulation at the level of the hundredths of milliseconds or the sets of ten Hertz (we cannot pronounce the same vowel exactly the same way every time, since speech organs are highly flexible).

Because of this, none of the two properties can independently identify particular productions of two different vowels, but they do represent a general tendency and can differentiate classes of vowels (standard deviation values are not shown, but they are generally very close to the difference between the average value of a given vowel and the average for all vowels, both for F0 and length).

Despite this variability, there are still patterns that can be observed at several levels: the segmental level (segment class, i.e. inherent variation), segmental context, syllabic context (vowel nasality should be included here if we follow the phonological analyses that propose that nasality is a property of syllable nucleus, which seems very adequate for E.P. – see [10]).

3.1.1. Intrinsic Variation

We can see that, when all other distinctive features are equal ([+/-round, +/-back]), the lower the vowel, the longer it will be, and the lower the fundamental frequency associated with it (e.g., in Table 1, [i] is 13% shorter than the average vowel whereas [ɛ] is 8% longer). This is in accordance to what has been found for other languages [14].

There are a few exceptions to the above generalization, but low vowels are consistently longer than high ones and are persistently associated with lower fundamental frequency.

In fact, for every given column in any table, the longest vowel – as well as the vowel displaying the lowest F0 value – is always a low one (or non-high if it is a nasalized vowel; recall that low nasal vowels can only be found dialectally).

Also, the shortest vowel – as well as the vowel associated with the highest F0 – is always a high one. Differences in F0 are more significant in female productions.

There does not seem to be any correlation between F0 or duration and the [+/-back, +/-round] features.

3.1.2. Segmental context

From Table 5, a few trends can be observed.

A vowel preceding a voiced obstruent is longer than the same vowel preceding a voiceless obstruent, which tallies with findings for other languages. Interestingly, a vowel after a (non-flap) voiced consonant is also longer than when it follows a voiceless consonant (or a flap). A vowel before a plosive is slightly shorter than a vowel before a fricative or a sonorant consonant. A vowel's duration is larger before a flap than before any other type of consonant.

3.1.3. Syllabic context

Vowels in closed syllables (Table 3) tend to be shorter than in open syllables (the difference is around 10%). It appears that this difference somewhat increases with vowels' height, but this is tentative.

If we compare Tables 1 and 2, we can see that vowel F0 does not appear to be affected by the closed or open nature of the syllable where that vowel occurs, except that it may not be as distinctive in closed syllables as it is in open ones.

Table 4 clearly shows that nasal vowels are much longer than oral ones (42% longer on average).

Finally, prevocalic glides are much longer than postvocalic semivowels (Table 6).

4. Discussion

It will be interesting to notice the differences between these data and some results obtained for other languages.

What is perhaps the most striking discrepancy is that, in E.P., F0 only seems to distinguish vowels and is actually not affected by the voiced/voiceless nature of the preceding consonant or its manner of articulation. Another noteworthy difference is that these factors have consequences on vowel length instead. The details are suggestive: whereas in other languages the state of the glottis during the production of a consonant influences the next vowel's F0 especially if that consonant is a plosive, in E.P. the effect is on vowel length rather than on F0 but is more salient with plosives, too. So there are reasons to suspect that this difference might be connected to some other aspect that would be worth looking at. In fact, [12] reports that the voiced/voiceless distinction has no effect on F0 in Brazilian Portuguese and, in accordance with other studies (e.g., [7]), the same author suggests that this cross-linguistic difference is due to a more fundamental difference (also related to the glottis), namely that languages where voiceless stops are aspirated are languages where these consonants influence F0, and languages where they are not aspirated are languages where they do not affect this property. This hypothesis has met a mixture of supporting arguments together with some counterexamples, and we shall not discuss it further.

Another peculiar fact that we mentioned is that vowels are longest before the alveolar flap (the uvular trill was left out of consideration because it is often produced as a fricative). The effect is of small magnitude, but nevertheless if we think of other languages where consonants of a similar nature produce lengthening effects on the previous vowel that are perceptively salient enough to deserve a phonological description, this is no doubt interesting, because E.P. is not considered to display this effect, despite the fact that, somewhat like those languages, in certain contexts the flap can have zero realization [11] (namely at the end of the word – but no words with this property are included in these results).

We have not presented duration values for semivowels adjacent to nasal vowels. We have not done so because they pattern the same way as glides next to oral vowels. One final point is that the different length of prevocalic and postvocalic glides that we observed is phonologically relevant, because semivowels deserve different phonological representations in terms of syllable structure according to their position relative to the neighboring vowel – see [10].

5. Conclusions

There are several factors that have been pointed out in the literature in order to explain the difference in inherent duration and intrinsic fundamental frequency across vowels.

For instance, it has been claimed that low vowels are longer than high ones because of the additional time required to open one's mouth when uttering them. Also, nasal vowels are also much longer purportedly because of the additional movements of the velum. F0 variation has also been related to tongue movement affecting the larynx [8].

One might also acknowledge that these differences are phonologically convenient, because they serve as an extra way to differentiate vowels, which become more stable and resistant to phonological fusion as they are also perceptually more distinct. This has also been suggested elsewhere (e.g. [5]).

Most of our data corroborate the patterns that have been found for other languages, supporting the hypothesis that many of these effects are indeed universal.

It should be noted that our results are to be considered with care. Although some of them are statistically significant, they are not necessarily important, since the observed differences are often of small magnitude.

6. Acknowledgements

We would like to express our gratitude to Eng. Onofre at the Centro de Análise e Processamento de Sinais of the Instituto Superior Técnico for all help and for making it possible for us to conduct recordings in the anechoic chamber in that center.

We would also like to acknowledge the contributions to this study by our colleague Eva Arim, who also worked on this experiment.

This study was part of project PRODIP, funded by FCT (Fundação para a Ciência e Tecnologia), with code POCTI/LIN/45341/2002.

7. References

[1] Delgado-Martins, M. R., 1973. Análise acústica das vogais tónicas do português. *Boletim de Filologia* 22(3), 303-314.

[2] Delgado-Martins, M. R., 1975a. Vogais e consoantes do português: estatística de ocorrências, duração e intensidade. *Boletim de Filologia* 24(1), 1-11.

[3] Delgado-Martins, M. R., 1975b. Perception de l'accent en Portugais. *Actas do 8º Congresso Internacional de Ciências Fonéticas*. Leeds.

[4] Di Cristo, A.; Hirst, D.J., 1996. Modelling French micromelody. *Phonetica* 43(1), 11-30.

[5] Diehl, R. L., 1991. The role of phonetics within the study of language. *Phonetica* 48, 120-134.

[6] Falé, I., 1997. Duração das vogais tónicas e fronteiras prosódicas: uma análise em estruturas coordenadas. *Actas do 13º Encontro Nacional da APL,* M. A. Mota & R. Marquilhas (eds.). Lisbon: Colibri, 255-269.

[7] Halle, M,; Stevens, K. N., 1971. A note on laryngeal features*. MIT QPR* 101, 198-213.

[8] Lehiste, I., 1970. Suprasegmentals. Cambridge, Massachusetts: MIT Press.

[9] Lehiste, I.; Peterson, G. E., 1962. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33, 419-425.

[10] Mateus, M. H. M.; Andrade, E., 2000. The phonology of Portuguese. Oxford: Oxford University Press.

[11] Mateus, M. H. M.; Rodrigues, C., (to appear). A vibrante em coda no Português Europeu. *Actas do XIX Encontro da Associação Portuguesa de Linguística*.

[12] Oliveira, M., 1998. A comparative study between American English and Brazilian Portuguese intrinsic pitch of vowels. *Proceedings of the 14th Northwest Linguistics Conference*, K. Lee & M. Oliveira (eds.). Vancouver: Simon Fraser University, 146-155.

[13] Viana, M. C., 1979. O índice duração e a análise acústica das oclusivas orais em português. *Boletim de Filologia* 25(1), 1-20.

[14] Whalen, D. H.; Levitt A. G., 1995. The universality of intrinsic F-0 of vowels. *Journal of Phonetics* 23(3), 349-366.