ACOUSTICAL CHARACTERISATION OF THE ACCENTED SYLLABLE IN PORTUGUESE; A CONTRIBUTION TO THE NATURALNESS OF SPEECH SYNTHESIS

*Teixeira, João Paulo – ESTiG-Bragança and CEFAT (F.E.U.Porto), e_mail:*joaopt@ipb.pt
Paulo, Elisabete Rosa- ESTiG-Bragança, e_mail: elis_paulo@hotmail.am
Freitas, Diamantino - CEFAT (F.E.U.Porto), e_mail: dfreitas@fe.up.pt
Pinto, Maria da Graça – F.L.U.Porto, e_mail: mgraca@fl.up.pt

ABSTRACT:

Text-to-Speech systems require control of the prosodic parameters of the produced speech waveform in order to achieve a higher naturalness and degree of perception. Amongst the several dimensions into which prosody can be unfolded, the accented syllable realisation brings the basic problem of producing a set of comprehensive rules for accurate control of the acoustic realisation of the syllable parameters, which remains to be solved. In particular, for the Portuguese language, in Europe, a set of comprehensive quantitative characterisation data and rules is absolutely lacking. The present paper is intended, as a quantitative contribution, as far as we know the first, to the solution of this problem. The duration Intensity, and variation of F0 were modelled in the tonic syllable according to its position in the word and the position of the word in the sentence.

1. INTRODUTION

With the objective of construction of prosodic models to improve the naturalness of synthetic speech it is recognised by some authors [1], [2], [3] that the correct modelling of the tonic syllable prosody has a crucial importance. Some authors agree in the modification of the acoustic parameters duration, intensity and F0, but there are no published works that quantify the variation of these parameters, at least for Portuguese.

The variation of F0, intensity or duration of the tonic syllable may depend on the meaning of the word in the context, or the length of the word, or the type of word according to the position of tonic syllable (oxiton, paroxiton, proparoxiton), the position of the tonic syllable or the position of this word in the sentence (in the beginning, middle or end).

The main motivation of this work concerns the utilisation of the results by TTS systems that generally may identify the position of syllables in words and the position of words in phrases.

Considering only the position dependence, the main objective is develop a model to implement the variations of F0, intensity and duration in the tonic syllable. In the study the categories defined by the positions were kept. .

2. METHOD

2.1 Corpus

A short corpus in which the tonic syllable always contains the phoneme [ε] was considered, bearing in mind that this study should be extended, in a second phase, to a larger corpus with other phonemes and with refinements in the method (reported below) resulting from this first phase.

Two words were considered for each position of tonic syllable (ex: **fe**rro, A**mé**lia, ca**fé**). Three sentences were created with each word, and a sentence with the word alone was also considered . The non sense word "fe**fe**to" was also included..

This is a total of 25 sentences. The characteristics of the tonic syllable were then extracted and analysed in comparison with a neighbour reference syllable (unstressed) in the same word. The non-sense word is full of interest because it contains the same syllable twice, in pre-tonic or post-tonic positions, allowing the reference syllable to be the same as the tonic syllable.

2.2. Recording Conditions

Each of the three speakers (2 males and 1 female) read each sentence 3 times in an auditorium with some acoustical treatment and the voice sound was recorded using a unidirectional microphone.

A resolution of 16 bits and a sampling frequency of 11 kHz were used.

2.3. Signal analysis

In the MatLab environment a tool was created to do the measures. The signal parts were first classified into voiced, unvoiced, mixed, and silence. In each voiced part the F0 contour was determined using cepstral analysis [4]. In all signal regions the intensity in dB was calculated according to Rowden [5]. The three graphic plots with the signal and classification, F0 contour and intensity were plotted on screen for inspection. The program allows the reproduction of parts of signals and zooms.

A set of measurements was registered in each tonic syllable and one of the neighbour syllables that is, in each word, used as a reference syllable.

The set of measurements comprehends the duration of the syllable, the maximum value of intensity and the initial and final F0 values of the syllable, as well as the contour of variation.

2.4 Data analysis

The data analysis was very simple and consisted in the determination of average values for each of the 3 speakers and the ensemble in each case of position of the stressed syllable in the word and position of the word in the phrase. A variance study was also done in order to determine the more consistent tendencies.

3. RESULTS

3.1 Duration

For each speaker the average percentage variation of the tonic syllable duration relative to the reference was determined and the tendencies were observed

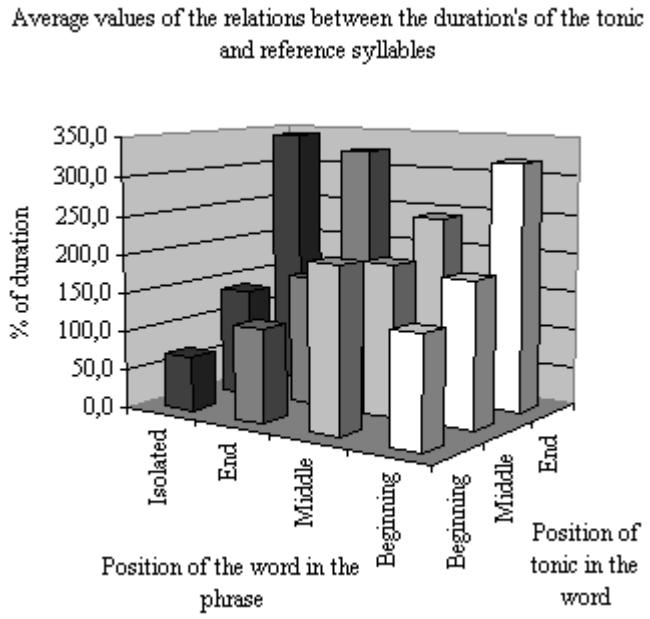

Figure 1

in function of the position of the tonic syllable in the word and the position of this word in the phrase. The variations pattern and the values are very close for the 3 speakers, leading us to the conclusion that the variation of the duration in the tonic syllable is basically independent of the speaker. Figure 1 represents the average values of the three speakers that has a low variance.

Some rules appear from the data and can be seen in the graphics, always relatively to the reference:

- for the words in the middle of the sentence, the duration of tonic syllable is approximately the double (200%).
- For other word positions this syllable will be:
  - About 300% if is in the end of word.
  - About 160% if is in the middle of word.
  - Between 120% and 140% if is in the beginning of word.

Some considerations have to be taken from these results. Firstly, the reference syllable is not the same as the tonic syllable. Secondly the results were obtained for a specific set of syllables. Will, for other syllables, the results be the same? Thirdly what is the meaning of longer syllables for synthesis. In a longer syllable which constituents are longer? only the vowel? also the consonant? How does it depend of the typ of consonant (stop, fricative, nasal, lateral)?

3.2 Intensity

For each speaker the average difference of intensity between tonic and reference syllables was determined in dB according to the position of the tonic syllable in the word and the position of this word in the phrase. For all speakers there are common patterns of decreasing intensity in the tonic syllable according to its position, from the beginning to the end of the word.

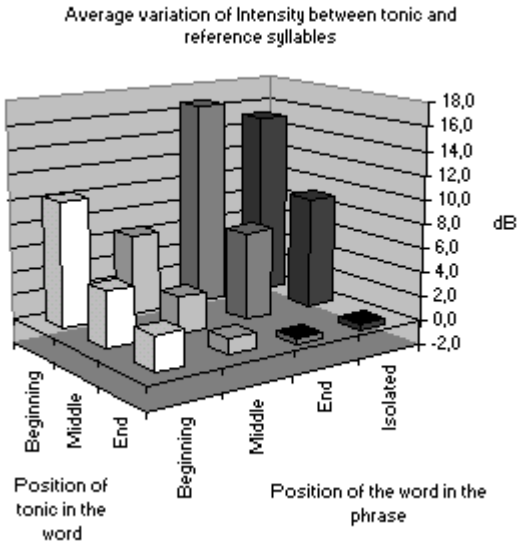

Figure 2

Figure 2 shows the average intensity of the tonic relative to the reference syllable. The major result is that the intensity is always higher in the tonic syllable. This increase is not so strong for the words in the middle of the phrase. The intensity increases in the tonic syllable depending of its position in the word as:

- 6 – 16 dB in the beginning of word.
- 3 – 9 dB in the middle of the word.
- 0 – 3 dB in the end of word.

The slope of variation of intensity of the tonic syllable in the word is noticeably higher in the last word of a phrase.

As in these experiments the tonic syllable always contains the phoneme [ε], that is one rather open phoneme and strongly pronounced, how much does this affect the results? In order to eliminate this problem the reference syllable should ideally be the same as the tonic, even if we have to use non sense words like ("fefeto").

3.3 Fundamental Frequency

The variation of F0 value in % (F0final/F0initial) was measured in the tonic and reference syllable. As theses syllables are in neighbour positions the common variation of F0 value is the result of sentence intonation. The difference of variations of F0 in these two syllables is then due to the tonic position. For each speaker the average difference of F0 relative variation between tonic and reference syllable [F0final/F0initial (tonic) - F0final/F0initial (reference)] was determined. There are some common trends of the results of the three speakers and some other variations that seam irrelevant. Figure 3 shows the results for the three speakers. The most relevant conclusions identifiable in the graphic are:

- when the tonic syllable is in the end position of the word the F0 value falls down about:
  - 10% if the word is in the end of phrase.
  - 20% if it is in a isolated word.
- If the word is in the beginning of the phrase the F0 value in the tonic syllable rises:
  - 5% if this syllable is in the beginning of word.
  - 10% if is in the end or middle of word.

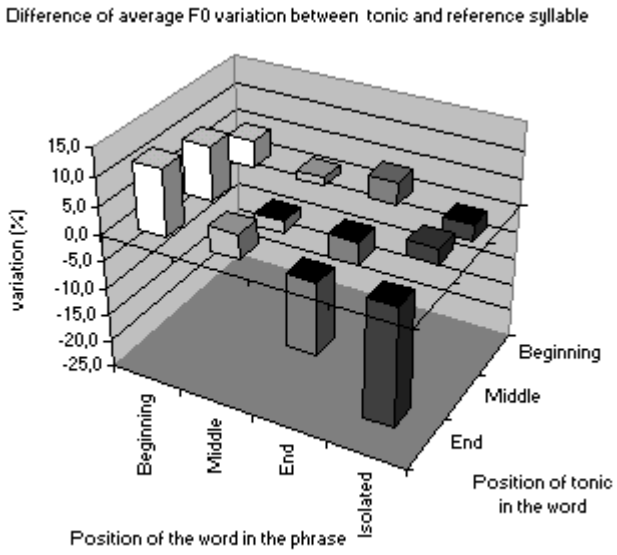

Figure 3

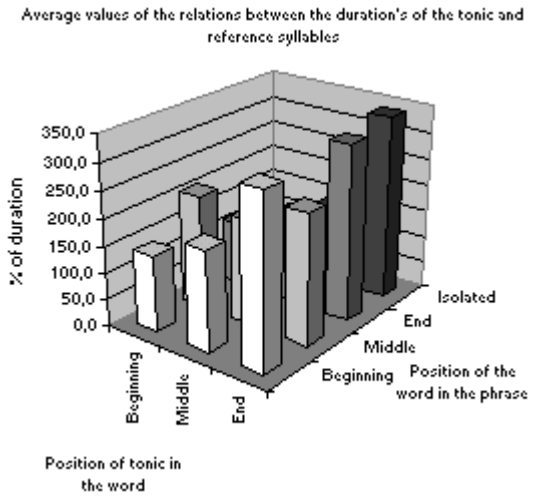

Figure 4.a

Only the values of variations are reported, but the patterns of variations are important as well. The curves were registered and in most of the cases the aspect resembles exponential functions.

The presented results are the variation of F0 related to its initial value. It's still necessary in practical synthesis situations to determine the appropriate F0 value at the start of tonic syllable.

4. CONCLUSION

Interesting opposite variations of duration and intensity in the tonic syllable according to its position in the word, for words in initial or final position in the phrase or isolated words. As shown in figures 4.a and 4.b for the position of tonic syllable along the word (beginning to end) the intensity decreases and the duration increases.

After this preliminary study some aspects about the method became clear. It's important to take them into account for further developments.

It is desirable that the reference syllable for comparisons of duration to be the same syllable as the tonic, but without context. It is also important to know what happens to the durations of the consonants.

For the intensity values, the reference syllable must be neighbour of the tonic, but should be the same syllable. This leads to use of non-sense words containing the same syllable twice.

5. FUTURE DEVELOPMENTS

We pretend to extend this study to other vowels.

For the intensity it is desirable to have a corpus with words that contain the same syllable as the reference.

For durations and F0 value a very large corpus of text should be used. In this corpus we would determine the default duration of each syllable and compare it to the duration of the same syllable when in tonic position.

The F0 variation in the tonic syllable is independent of the syllable. The reference syllable must be the

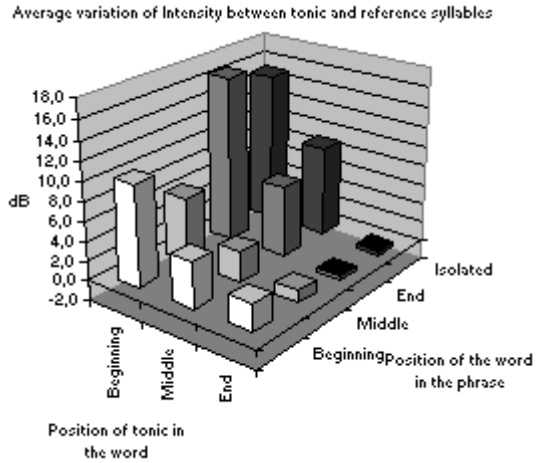

Figure 4.b

neighbour syllable. We just need more cases to validate our results.

6. REFERENCES

[1] Zellner, B., 1998. Caractérisation et prédiction du débit de parole en français, Thése Doct. Lausanne.

[2] Andrade, E. e Viana, M. C., 1988. Ainda sobre o ritmo e o Acento em Português. In Actas do 4º

Encontro da Associação Portuguesa de Linguística, Lisboa.
[3] Mateus, M. H. M., Andrade, A., Viana, M. C. E Villalva, A. , 1990. Fonética, Fonologia e Morfologia do Português. Universidade Aberta.
[4] Teixeira, J. P. , 1995. Modelização Paramétrica de Sinais para Aplicação em Sistemas de Conversão Texto-Fala. Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto.
[5] Rowden, C. , 1992. Speech Processing. McGRAW-HILL Book Company.