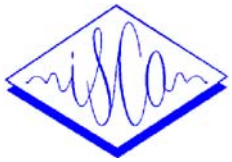
ISCA Archive
http://www.isca-speech.org/archive



EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

# CATALAN VOWEL DURATION

*Lourdes Aguilar, Julia A. Giménez, Maria Machuca, Rafael Marín, Montse Riera*
Departament de Filologia Espanyola, Universitat Autònoma de Barcelona,
Facultat de Lletres, Edifici B, 08193, Bellaterra, Barcelona, Spain
{lourdes, aliju, maria, rafa, montse}@liceu.uab.es

## ABSTRACT

The temporal organization of discourse has produced a great deal of works in several languages pointing to different aims: from studies where the identification of cues about the planning of linguistic message is treated to studies in which duration models for text-to-speech systems are proposed.

This work is a first step towards the description of Catalan vowel duration. Considering the Catalan vowel system, two subsystems can be distinguished according to stress: stressed vowels: /i/, /e/, /E/, /a/, /O/, /o/, /u/, and unstressed vowels /i/, /u/, /@/. The purpose of the present study is to provide data for Catalan vowels in order to achieve a data-oriented description and at the same time a predictive model suitable to be implemented in a TTS system.

10.21437/Eurospeech.1997-261

## 1. INTRODUCTION

Due to the fact that the control of prosodic parameters is considered one of the main problems to obtain high quality text-to-speech (TTS) systems, the results of experiments on segmental duration are usually used to develop duration models which can predict in an accurate way the temporal properties of speech synthesis units ([1], [2], [3], [4], [5]). In order to improve the naturalness of TTS systems, it is acknowledged that data referred to vowels are more determinant than those ones referred to consonants. From a descriptive approach, this can be related to the greater number of factors affecting vowel duration.

The purpose of this study is to provide data for Catalan vowels from an experimental analysis where a number of factors is taken into account not only to obtain a description for each vowel, but also to propose a duration model which could be included in a TTS system.

Considering the Catalan vowel system, two subsystems can be distinguished: the stressed one, including /i, e, E, a, O, o, u/, and the unstressed one, with /i, u, @/ (SAMPA transcription is being used [6]). Vowel reduction in an unstressed syllable is phonologically relevant: in unstressed positions, vowels [a, e, E] are reduced to the central vowel [@], whereas [O, o, u] are realized as [u] (see [7] for a more detailed description).

## 2. EXPERIMENTAL PROCEDURE

The experimental design consists in comparing the acoustic manifestation of Catalan vowels in read texts according to a set of specified variables. To acquire the samples of speech, words containing the items of interest have been incorporated in a text, which has been read by three male speakers.

The corpus was constructed attending to the following variables: sentence position, stress, syllable type, and voicing and manner of articulation of post-vocalic consonant. Other factors such as word grammatical function, stress pattern and place of articulation of the post-vocalic consonant were controlled. In total, the corpus is composed of 556 trisyllabic content words.

Items of the corpus were analyzed by means of the speech analysis software MacSpeechLab in a Macintosh II. Waveform displays and broad-band spectrograms were plotted for each sequence, and duration measurements were taken. The total number of analyzed items is 1668 (556 items x 3 speakers).

## 3. RESULTS

Data obtained in the experimental analysis are organized so as to observe the effects of the considered factors on vowel duration. Afterwards, these effects are presented as descriptive rules.

### 3.1. Factors affecting vowel duration

Since Catalan has a stressed vowel system and an unstressed one, an individual treatment for each system has been performed. We can say, however, that besides the phonological factors that favors this treatment, acoustic differences arise.

As it has been said before, vowels that can be found in a stressed position are not the same than the ones which appear in an unstressed position. Therefore, inter-vowel comparisons depending on the presence of stress were only done for /i/ and /u/: pooling stressed /i/ and /u/, a mean duration value of 85.01 ms is obtained, while pooling unstressed /i/ and /u/, a mean duration of 73.64 ms is found out. Differences are significant, as pointed by an ANOVA test (p:.0001).

#### *3.1.1. Sentence position*

As it can be observed in Figure 1, vowels in prepausal position (x=113.2 ms) show invariably a longer duration than vowels in non-prepausal position (x=83.5 ms). The sentence position effect is independent on the stressed or unstressed nature of the system, as pointed out by a two-way ANOVA test where significant differences arise, both due to the effect of sentence position (p:.0001) and to the effect of stress (p:.0001).

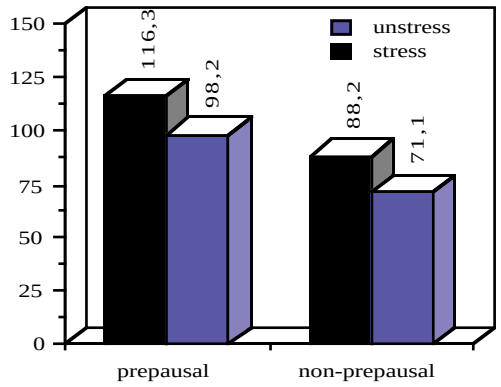



Figure 1. *Mean vowel duration values depending on sentence position and stress.*

*3.1.2. Voicing of post-vocalic consonant*

In addition to the effect of sentence position, an influence of post-vocalic consonant can be described, although only as far as non-prepausal position is concerned, due to corpus restrictions.

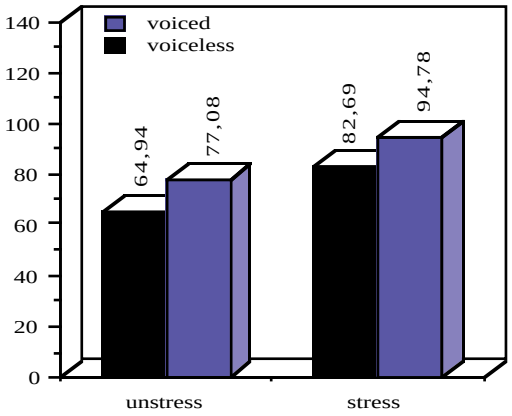



Figure 2. *Effect of the voicing of post-vocalic consonant on duration values in stressed and unstressed systems, in non-prepausal position.*

The voicing of the consonant following the vowel causes an increase of its duration: in non-final position, stressed vowels appearing before a voiced consonant are longer than the corresponding vowels before voiceless consonants, as shown in Figure 2. An ANOVA test points out significant differences in the stressed system (p:.0001) as well as in the unstressed one (p:.0001).

*3.1.3. Manner of articulation of post-vocalic consonant*

As regards to the manner of the consonant following the vowel, a lengthening effect has been found in prepausal position: vowels preceding fricative consonants are longer than those preceding plosive ones (a mean of 121.29 ms compared to a mean of 103.52 ms). These differences are significant, as inferred from an ANOVA test (p:.0001).

With respect to non-prepausal position, when the consonant following the vowel is voiceless, a comparison between stops and fricatives shows a relevant difference (a mean of 75.47 ms compared to a mean of 80.49 ms, respectively; p:.0001); if the consonant following the vowel is voiced, a comparison between approximants and voiced fricatives reproduces the effect (85.13 ms and 91.96 ms, respectively; p:.0001).

*3.1.4. Syllable structure*

On the contrary, syllable structure does not appear as a determinant factor in Catalan vowel duration.

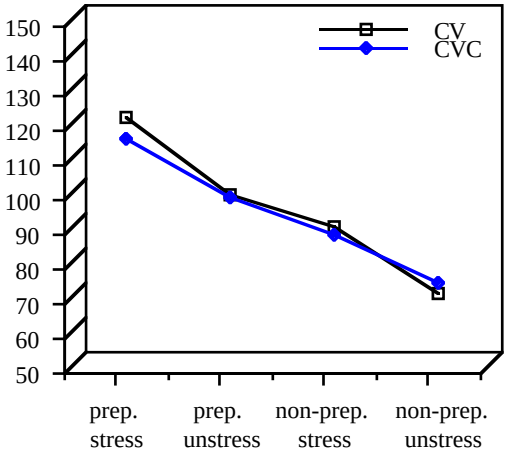



Figure 3. *Effect of syllable structure on mean duration values in stressed and unstressed systems, in prepausal and non-prepausal positions.*

As shown in Figure 3, there is a slight tendency to increase vowel durations in CV syllables compared to CVC syllables, but this does not arise as significant in a two-way ANOVA test (stress x syllable structure) for either prepausal (p:.35) or non-prepausal position (p:.58).

*3.1.5. Vowel identity*

Besides the study of factors affecting duration values, it is worth deciding if duration can be considered as a cue of vowel identity.

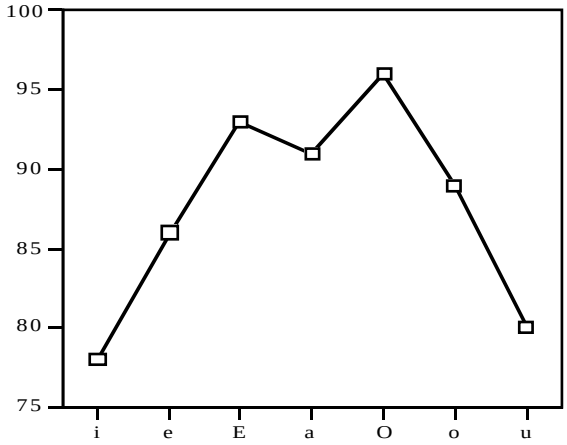



Figure 4. *Mean duration values of stressed vowels in non-prepausal position.*

With respect to data concerning stressed vowels in non-prepausal position, a gradation O > E> a > o > e > u > i is found, as it can be observed in Figure 4, while a gradation @ > u > i is revealed for the unstressed system (72, 71, 70 ms. respectively).

However, these differences are only significant in the stressed system (p:.0001). Related to this, a Scheffe-test, that locates differences between stressed vowels, allows us to organize vowels in three groups along the open-close dimension: a/E/O, e/o, i/u.

As for prepausal position, an arrangement E=a > e=O > i > o > u in the stressed system and a i> u > @ in the unstressed system can be described. Nevertheless, differences are not significant in any case, as inferred from an ANOVA test (p:.12 and p:.33, respectively).

| | prepausal | | |
|---|---|---|---|
| | n | x | sd |
| **i** | 30 | 111.8 | 25.2 |
| **e** | 27 | 119.6 | 22.7 |
| **E** | 28 | 126.3 | 22.9 |
| **a** | 56 | 126.3 | 28.0 |
| **O** | 27 | 119.9 | 26.7 |
| **o** | 28 | 110.7 | 22.6 |
| **u** | 25 | 108.8 | 29.6 |

Table I. *Number of cases (n), mean values (x) and standard deviation (sd) of the duration values of stressed vowels in prepausal position.*

| | n | x | sd |
|---|---|---|---|
| **i** | 12 | 105.6 | 5.4 |
| **@** | 19 | 94.3 | 4.09 |
| **u** | 13 | 97.2 | 6.97 |

Table II. *Number of cases (n), mean values (x) and standard deviation (sd) of the duration values of unstressed vowels in prepausal position.*

### 3.2. Some duration rules

From the results obtained, some duration rules can be formulated:

a) A vowel in a prepausal position is lengthened 35.56% with respect to a vowel in a non-prepausal position.

b) A non-prepausal vowel preceding a voiced consonant is lengthened 14.53% with respect to a non-prepausal vowel preceding a voiceless consonant.

c) A non-prepausal vowel preceding a voiceless fricative consonant is lengthened 6.65% with respect to a non-prepausal vowel preceding a voiceless stop consonant.

d) A non-prepausal vowel preceding a voiced fricative consonant is lengthened 6.83% with respect to a non-prepausal vowel preceding an approximant consonant.

e) A prepausal vowel preceding a voiceless fricative consonant is lengthened 17.16% with respect to a prepausal vowel preceding a voiceless stop consonant.

f) A non-prepausal vowel /i/ and /u/ in a stressed context is lengthened 12.5% with respect to the non-prepausal vowel in an unstressed context.

g) A prepausal /i/ and /u/ in a stressed context is lengthened 8.9% with respect to a prepausal /i/ and /u/ in an unstressed context.

## 4. MODELING VOWEL DURATION

The study of the effects of a core of specified variables on Catalan vowel duration and the establishment of some descriptive rules is the first step towards the design of a duration model suitable to be implemented in a TTS system.

The observed data make us thinking that a multiplicative model is more adequate than an additive model to describe the effects of the variables on a given reference duration , since the modification due to these variables does not appear in absolute, but in relative basis ([8]).

On the other hand, any interaction effect of the variables has been found and the effects related to the considered variables do not sum up; in other words, a joint independence cannot be inferred. Our data fit better to the incompressibility model proposed by [3]: each rule cannot act independently because a vowel reduced by the application of one rule tends to resist additional shortening.

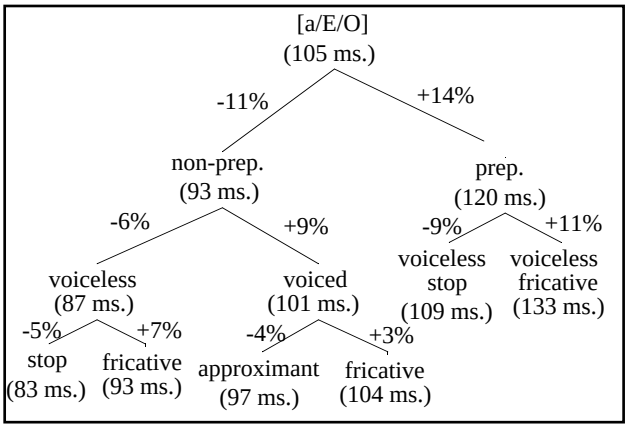



Figure 5. *Binary tree corresponding to /a, E, O/ duration assignment.*

To illustrate this, Figure 5 presents the model corresponding to stressed vowels /a, E, O/. The same should be done for /e, o/ and /i, u/, thus completing the stressed system. By means of the separation of stressed and unstressed systems, and assuming that vowels can be grouped according to their open- close properties, a total number of five trees will be needed: three trees in stressed system plus two trees corresponding to unstressed system (for /i, u/ and for /@/).

Nevertheless, an alternative option to decrease the number of trees could be covering by a single model vowels /i, u/ appearing in both stressed and unstressed systems, as shown in Figure 6.

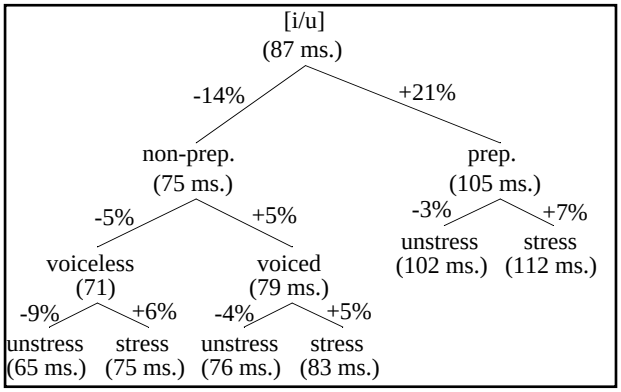



Figure 6. *Binary tree corresponding to /i, u/ duration assignment.*

The final decision is going to depend on the structure of the system in which duration assignment will be included.

## 5. DISCUSSION AND CONCLUSIONS

In this paper, several questions around Catalan vowel duration have been addressed to the data obtained from the analysis of a corpus of read text. Firstly, the effect of a set of specified variables has been analyzed; and then, a first attempt to the construction of a duration model has been presented.

As far as the effect of the considered variables is concerned, evidences also found for other languages including English, German, Danish, Swedish and Spanish have been reproduced in this experiment ([9], [10]). The well-known effect of prepausal lengthening is found independently on any other effect ([3], [8]). In addition, the effect of stress on vowel duration has been manifested, according to results for Catalan presented in [11]: on the one hand, vowels belonging to the stressed system are globally longer than vowels belonging to the unstressed one; on the other hand, comparisons for /i/ and /u/ depending on the presence of stress show this effect.

Finally, an important effect due to the voicing and manner of post-vocalic consonant has appeared in the treatment of the corpus. With respect to voicing, vowels preceding voiced consonants are always longer than vowels preceding voiceless ones; as for the manner, a lengthening effect of fricatives and approximants, when compared to stops, on vowel duration has been found. Evidence for this has also been described for several languages ([3], [12]). As for the existence of an intrinsic vowel duration, it can be observed that high vowels are the shortest in both stressed and unstressed systems, showing a correlation between tongue height and vowel duration. These results are in agreement with studies done for Catalan [7] and other languages [9].

It has been proved that data obtained in the experimental analysis can be used to build a model that predicts Catalan vowel duration. It seems that a multiplicative model fits the observed data.

Nevertheless, some questions remain open, such as the need of investigating some other factors (for instance, the grammatical function of the word) or the testing of the model in a TTS system.

## ACKNOWLEDGEMENT

R. Marín's work has been partially supported by a grant from *Generalitat de Catalunya* (AP-LP/95-2704).

## REFERENCES

[1] Umeda, N. (1975). "Vowel duration in American English", *Journal of the Acoustical Society of America*, 58 (2): 434-445.

[2] Crystal, T. H. and A. S. House (1988). "The duration of American-English vowels: an overview", *Journal of Phonetics*, 16: 263-284.

[3] Klatt, D. H. (1973). "Interaction between two factors that influence vowel duration", *Journal of the Acoustical Society of America*, 54: 1102-1114.

[4] Macarrón, A., G. Escalada and M. A. Rodríguez (1991). "Generation of duration rules for Spanish text to speech synthesis", *Eurospeech'91*, Geneve, Italy, 617-620.

[5] Marín, R. (1995). "La duración vocálica en español", *Estudios de Lingüística*, 10: 213-226

[6] Wells, J. C., W. Barry, M. Grice, A. Fourcin, and D. Gibbon (1992). "Standard Computer-Compatible Transcription. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992" in ESPRIT PROJECT 2589 (SAM) *Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report. Year Three: 1.III.91-28.II.1992*. London: University College London.

[7] Recasens, D. (1991). *Fonètica descriptiva del català*, Institut d'Estudis Catalans, Barcelona: Biblioteca Filològica XXI.

[8] van Santen, J. (1992). "Deriving text-to-speech durations from natural speech" in G. Bailly, C. Benoit and T. R. Sawalis (Ed.) *Talking machines: Theories, Models, and Designs*, Amsterdam: Elsevier Science.

[9] Lehiste, I. (1996). "Suprasegmental features of speech", in N. Lass (Ed.)*Principles of Experimental Phonetics*, Mosby.

[10] Borzone, A. M. and A. Signorini (1983). "Segmental Duration and Rythm in Spanish", *Journal of Phonetics*, 11: 117-128.

[11] Cantín, M. and A. Ríos (1991). "Análisis experimental del ritmo de la lengua catalana", *Anuario del Seminario de Filologia Vasca "Julio de Urquijo"*, XXV-2: 487-513.

[12] Laeufer, C. (1992). "Patterns of voicing-conditioned vowel duration in French and English", *Journal of Phonetics*, 16: 327-338.