Placing pauses in read spoken Spanish: a model and an algorithm

RAFAEL MARÍN, LOURDES AGUILAR, DAVID CASACUBERTA
Universitat Autònoma de Barcelona

Abstract

The purpose of this work is to describe the appearance and location of typographically unmarked pauses in any Spanish text to be read. An experiment is designed to derive pause location from natural speech: results show that Intonation Group length constraints guide the appearance of pauses, which are placed depending on syntactic information. Then, a rule-based algorithm is developed to automatically place pauses whose performance is tested by means of qualitative tests. The evaluation shows that the system adequately places pauses in read texts, since it predicts 81% of orthographically unmarked pauses; when pauses associated to punctuation signs are included, the percentage of correct prediction increases to 92%.

1. Introduction

Prosodic modeling is a research area in which both efforts to automate data analysis and to develop computational models are equally needed, given the complexity of the task. In particular, to overcome the lack of naturalness in speech generated by TTS systems, an adequate prosodic segmentation becomes crucial, due to its consequences in message understanding and acceptability (Nooteboom *et al.,* 1978).

The division in major and minor prosodic groups is the result of a combination of pauses and F0 movements, but rules governing their appearance are not easily reduced to one single factor; a good review of different factors affecting phrasing is given in Cruttenden (1986). In our opinion, the most useful way to deal with the problem of incorporating these factors into an automatic system is by providing a model in which a complete account of units, levels, and interaction between levels is given.

In the present study, we will focus on the appearance and location of pauses in natural speech to build a model suitable to be incorporated in a TTS system for Spanish. The problem of assigning pauses in TTS systems has been tackled in different ways, mainly due to the fact that treating prosodic segmentation involves decisions concerning the relationships between fields such as syntax, morphology, phonology and phonetics. Related to this, several questions arise: which of these disciplines can better account for pause assignment? Is it necessary to establish independent levels and design an interface between them? Is it preferable to consider different kinds of linguistic factors simultaneously? Solutions differ from systems performing a complete syntactic analysis, which is then reinterpreted in prosodic terms (Frenkenberger *et al.*, 1994), to algorithms that merely do a morphological analysis without considering syntactic information (Emerard *et al.*, 1992). The later procedure has been the preferred one in Spanish TTS systems so far (López, 1993; Castejón *et al.*, 1994). For Telefonica's TTS system, Castejón *et al.* (1994) propose a pause assignment method entirely based on part-of-speech labeling. Their module examines word categories so as to decide the segmentation of a sentence. First, punctuation marks -which are included as a category- are always associated to a pause; afterwards, if the resulting sequence is too long, additional pauses are located, according to morphological decisions. For instance, certain units such as coordinating conjunctions, subordinating conjunctions, verbs or function words favor the appearance of a pause.

Also for Spanish, assuming that there are correlations between prosodic and syntactic units, López (1993) develops an algorithm for prosodic segmentation based on coefficients which indicate the degree of syntactic cohesion between two adjacent words. If this degree is low, possibilities for a pause to appear increase; inversely, if the cohesion is strong, pauses are prohibited. This model can be seen as an attempt to formalize syntactic properties in terms of linear relationships, relying on the approach in Vergne (1993)[1].

Other descriptions, instead of treating linguistic levels separately, include syntactic properties, sentence length or discourse function of the words at the same time, because it is thought that all these factors are equally important (Ladd, 1987, 1996; Bachenko and Fitzpatrick, 1990; Hirschberg, 1991; Monaghan, 1992; Quené and Kager, 1992; Gili-Fivela

[1] This approach has also been applied in G. Vannier, A. Lacheret-Dujour, J. Vergne (1999). The text-to-speech system is in demo on: http://www.crisco.unicaen.fr/KaliDemo.html

and Quazza, 1997). If this viewpoint is adopted, a complete syntactic analysis is not necessary since syntactic features can be overruled by other types of linguistic information. For instance, the rules proposed by Bachenko and Fitzpatrick (1990) only have to examine a subgroup of the output coming from a syntactic parser; that is, access to lexical categories, nucleus placement or distribution of constituents is needed, but not to predicate-argument relations or modifier attachment. Likewise, O'Shaughnessy (1990) affirms that, for TTS systems, it is unnecessary to fully parse the text to be spoken: knowledge of verb location, major syntactic boundaries and stressed words is enough to achieve a proper generation of pauses.

Generally, it is the use of phonological domains such as the phonological phrase which allows the interaction between word level and syntactic constituents (Nespor and Vogel, 1986; Sosa, 1999). A good illustration of this is the work described in Hirschberg and Prieto (1996), who have adopted Pierrehumbert's intonational description for English to develop a phrasing module for a Mexican Spanish TTS system (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986). They also incorporate the advantages of automatic procedures, since rules are acquired from annotated text using Classsification and Regression Tree techniques.

In the present study, a model for pause assignment in Spanish is proposed, which relies on the use of a unit that conveys syntactic and prosodic properties, and relates morphological and phonological levels in a direct way. Section 2 is devoted to the description of text treatment procedure. In section 3, an experimental analysis is carried out to derive pause location from natural speech; results serve to develop an algorithm to automatically assign pauses in unrestricted text (section 4) whose performance is tested in section 5. Section 6 discusses the obtained results comparing them with recent works in the literature, and summarizes conclusions.

2. Units

Intonation group, stress group and categorial stress group are the units used in the development of the system. With respect intonation group (IG), although other criteria can be used to demarcate it (for instance, Cruttenden, 1986, refers to changes of pitch level or direction, and final syllable lengthening), here it is defined as the chunk of utterance between two pauses. In other studies, they have been called sense-groups, breath-groups or intonational phrases. Intonation groups correspond typically with major grammatical constituents like simple sentences, noun-phrase subjects or predicates.

In our model, intonation groups can be divided into stress groups (SG). The stress group is conceived as a string formed by zero or more unstressed words preceding a lexically stressed word. For the purpose of identifying SGs automatically, the term stress is used in a general way to mean prominence, saying nothing about the phonetic realisation of stress or the types of accent which occur in an overall intonation contour. Every word has one stress in its citation form, but some type of words most commonly occur in an unstressed form in connected speech. Related to this, it is assumed here that open categories are stressed in connected speech while closed categories are not.

An IG segmentation into SGs is illustrated in (1):

(1) [Le rogamos] [vuelva] [a marcar] [el número] [pasados] [unos minutos]
"Please dial the number after some minutes"

A categorial stress group (CSG) is an SG with added information about syntactic properties of the elements integrating it, represented by the lexical category of the first element in the group[2]. CSGs can be formed by only a stressed word, or by one or more unstressed words preceding the stressed one. As for the first CSG type, we simply indicate the syntactic category to which the word belongs. With respect to the second CSG type, the categories of the first and the last word are considered.

The list of open categories and the corresponding CSGs (between brackets) is as follows: adjective (ag), adverb (adg), gerund (gg), infinitive (ig), noun (ng), participle (ptg) and verb (vg).

The list of CSGs formed by more than one word is the following:

[2] A more detailed description can be found in Casacuberta *et al.*, (1998).

cg: conjunction preceded by adjective (a), adverb (ad), gerund (g), infinitive (i), noun (n), participle (pt) or v

ccg: coordinating conjunction preceded by adjective (a), adverb (ad),), gerund (g), infinitive (i), noun (n), participle (pt) or verb (v)

clg: clitic preceded by verb (v)

pg: preposition preceded by adjective (a), adverb (ad), gerund (g), infinitive (i), noun (n), participle (pt) or verb (v)

qg: quantifier preceded by adjective (a), adverb (ad), participle (pt) or verb (v).

The main advantage of CSG is that it is an enriched unit since it allows to incorporate syntactic features to the prosodic unit, SG: in other words, CSG allows to label sentences incorporating both prosodic features, because there is an identity between CSGs and SGs, and syntactic ones, since each CSG type has particular properties according to the categories of the words composing it. The sentence in (1) is labeled as shown in (2):

(2) [Le rogamos]$_{clg:v}$ [vuelva]$_{vg}$ [a marcar]$_{pg:i}$ [el número]$_{qg:n}$ [pasados] $_{ptg}$ [unos minutos]$_{qg:n}$

3. Experimental analysis

The aim of the experimental analysis is to describe the appearance and location of orthographically unmarked pauses in read texts in Spanish. The corpus consisted of 13 texts (1232 sentences) read aloud by a speaker. The recordings were marked with respect to pauses, defined as a perceptible silence, by a phonetically-trained subject, who listened to the recordings and used waveforms (obtained with the speech analysis software Waves +) to verify his auditive impressions. Independently, sentences in texts were segmented into CSGs[3].

From these data (location of pauses in read texts and CSG labelling), two kinds of analysis were performed in order to know, on the

[3] Segmentation was manual to avoid possible tagging problems.

one hand, if length constraints (measured in number of SGs[4]) guide the appearance of pauses, and on the other hand, if location of pauses can be related to the syntactic information offered by the CSG.

3.1 Effect of Intonation Group Length

The aim of the statistical treatment was to find out if sentence length, measured in number of SGs, affects pausing when punctuation signs are absent. The analysis is based on chunks of text between two punctuation signs with more than one CSG.

The relationship of Intonation Group length and pauses was computed: results are summarized in Figure 1, where the percentages of appearance of IGs are depicted as a function of the number of SGs integrating them. Percentages –in ordinate- are computed from the total number of IGs in texts. More clearly, the first column says us that 2% of the total number of IGs in texts has 2 SGs, the second column, that 14% % of the total number of IGs in texts has 3 SGs, and so forth.

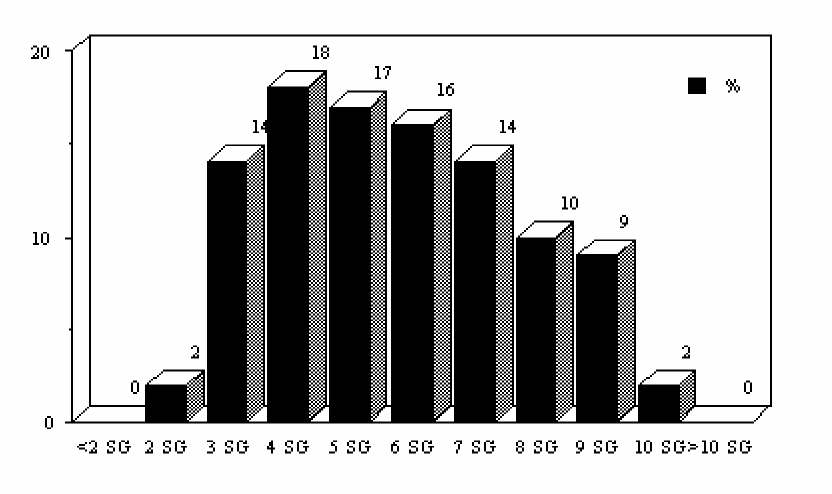

Figure 1. *Percentages of appearance of IGs depending on the number of SGs in IG.*

[4] Since there is an identity between CSG and SG, this metrics, instead of number of syllables, simplifies our model.

Figure 1 shows that the upward length constraint is more determinant than the minimal length constraint: preferences on the number of SGs in IGs range from 3 to 6. Intonation groups with only 1 SG are not present in corpus —as regards to non orthographically marked pauses—, as well as intonation groups with more than 7 SGs.

Despite the fact of the speaker´s choice involved in pause assignment, IG length constraints on placing pauses in read texts can be formalized:

1) There is an upward length constraint: The appearance of a pause is mandatory in sentences containing 10 or more SGs.

2) The tendency to locate the pause, when appearing, in a centered position (described in Nespor and Vogel, 1984) is formalized in terms of the distance in CSGs from the pause to the beginning and end of the sentence as well as in terms of the distance in CSGs from the preceding and following pauses.

A pause cannot generate an IG with less than 3 SGs, which implies that:

2a) A sentence with less than 6 SGs cannot be segmented.

2b) A pause cannot appear before 3 SGs of the beginning of a sentence, and before 3 SGs of its end.

2c) The minimal interpause distance cannot be inferior to 3 SGs.

Between sentences that have to be segmented and sentences that cannot be segmented, there is a range in which we can talk of optionality. In other words, if there is a minimal (6 SGs) and an upward (9 SGs) length constraint, it is reasonable to claim that pauses are optional in sentences between 6 and 9 SGs and that the appearance of a pause in them will be explained by other type of information. To know this, experiment 2 described in section 3.2. has been carried out.

3.2. Effect of grammatical information

In Spanish descriptions (Canellada and Madsen, 1987, Quilis, 1993), syntactic dependency on pausing has been explained by means of a set of combination of categories between which no pause can appear, for instance: adjective-noun, noun-adjective, verb-adverb, adverb-verb, adverb-adjective, adverb-adverb, determinant-noun, determinant-adjective, elements in colocations.

Many of these forbidden boundaries such as preposition/noun, article/adjective, clitic/verb, conjunction/verb do not need to be implemented as rules in the system we are describing because a pause cannot appear inside an CSG, so any pause between an unstressed element and a stressed one, or between two unstressed ones is blocked. The problem arises whith the combination of two stressed elements. Besides this, different places in a sentence for a pause are usually possible, without changing meaning.

In order to find general trends governing the location of pauses, it was hypothesized that certain CSGs favor the appearance of a pause before them, where other difficult it. To verify this, the relative frequency of pause location preceding a CSG was calculated in the corpus data: the number of CSGs of a syntactic type followed by a pause was divided by the total number of this type of CSG found in the text.

type of CSG	total number of CSGs	number of CSGs preceded by pause	%
ccg	223	73	32
Vg	275	71	26
clg	62	14	22
Cg	199	34	17
Gg	23	4	17
adg	144	19	13
ptg	115	7	6
Pg	807	45	5
Qg	347	13	4
Ng	286	5	2
Ag	247	2	0,8
Ig	56	0	0

Table I. *Frequency of appearance of a pause as a function of the type of the following CSG.*

Table I presents results: for each CSG type (encoded according to paragraph 2), total number, number of pauses found before it and frequency of appearance are shown.

A hierarchy of CSGs depending on the index of probability of having a pause before it is derived from the table: ccg, vg, clg, cg, gg, adg, ptg, pg, qg, ng, ag, ig.

Hierarchy:

Place pauses according to this preference order: ccg, vg, clg, cg, gg, adg, ptg, pg, qg, ng, ag, ig.

The hierarchy is implemented in the system, in order to locate pauses in texts. In a given sentence, categories at the head of the list are the best candidates to have a pause before them. However, this hierarchy is not enough to prevent certain prohibited pauses, such noun and adjective. To solve this, besides the probabilities list, a set of rules derived from the literature are used to check if pause is being to appear in a banned position.

Restriction rules:

Block pauses between ag-ng, ng-ag, vg-adg, adg-vg, adg-ag, ag-adg, adg-adg.

4. The System

From the empirical results, a tool (henceforth, ProPause) has been developed to locate orthographically unmarked pauses in Spanish texts, with the aim of validating the model. Procedures are applied once orthographically marked pauses have been discarded: in other words, they work on chunks of text between two punctuation signs. A main module loads all the subprograms needed, asks for a file to be segmented into pausal units, and processes the text within the file using the different ProPause modules: CSG Categorizer, CSG Counter and Pause Searcher.

4.1. CSG Categorizer

The CSG Categorizer automatically divides texts into CSGs from texts morphologically labeled. First, the program parses the text checking punctuation signs, and afterwards the resulting sequences are segmented into CSGs using stress information associated to each category. Namely, c, cc, cl, p and q are unstressed, while a, ad, g, i, n, pt and v are stressed.

4.2. CSG Counter

By applying the length constraints stated in section 3.1, this module decides whether an utterance has to be segmented or not. If the number of CSGs is less than 6, no pause is allowed; therefore, no more functions are invoked and the program shifts to the next utterance. On the contrary, if the number of CSGs is greater than 10, the pause is mandatory, so it is required to load the hierarchy containing all the CSG types ordered according to their probability to present a pause before them.

Finally, if the number of CSGs falls between 6 and 10 the pause is considered to be optional. In this case, the subset of the hierarchy that only refers to CSGs having a high probability for a pause to appear is loaded.

4.3. Pause Searcher

Once the previous module has decided the mandatory or optional nature of a pause, the Pause Searcher looks for the appropriate place to locate it. First, it applies the criterion stating that a pause cannot appear at a distance less than 3 SGs from the beginning and the end of the sentence; after this, the CSG hierarchy is used to find the most suitable place for the pause (cf. Paragraph 3).

Once a candidate for a pause is selected, the module checks if restriction rules prevent the breaking (cf. Paragraph 3). If this happens, backtracking is applied and the program proposes the next member in the CSG hierarchy. The process is recursively invoked until the utterance is divided into IGs that cannot be further segmented.

4.4. An example

To illustrate the procedure, a sentence is processed along this section. To start with, since the sentence in (3) has 12 SGs, the system has to put a pause, because of prosodic requirements.

(3) [El péndulo]qg [había]vg [comenzado]ptg [entonces]adg [su oscilación]qg [y la quietud]ccg [que reinaba]cg [entre nosotros]pg [era]vg [absoluta]ag [en el silencio]pg [de la noche]pg

'The pendulum had started then its oscillation and the calm prevailing among us was complete in the silence of the night'

Once the appearance of the pause is decided, it is needed to determine where it should appear. In order to model the prosodic constraint related to the distance of the pause from the beginning and the end of the sentence, the first 3 SGs and the last 3 SGs are discarded, obtaining the fragment in (4):

(4) [entonces] [su oscilación] [y la quietud] [que reinaba] [entre nosotros] [era]

Now, Propause uses the CSG hierarchy to find the best place for locating the pause, according to which it is preferred to put the pause before a [ccg], obtaining then the segmentation in (5):

(5) a. [El péndulo] [había] [comenzado] [entonces] [su oscilación]
b. [y la quietud] [que reinaba] [entre nosotros] [era] [absoluta] [en el silencio] [de la noche]

Nevertheless, before determining the definitive location of the pause, restriction rules are checked in order to avoid splitting some combinations of CSGs: as for our example, the phrasing is considered to be appropriate.

Once the first pause is assigned, the program examines if it is possible to divide any of the resulting sequences. It is the case for this sentence, because its second segment has 7 SGs, which makes the pause optional: since [vg] appears inside and no restriction rule applies, the segment is divided again, giving the two sequences in (6a) and (6b):

(6) a. [y la quietud] [que reinaba] [entre nosotros]
b. [era] [absoluta] [en el silencio] [de la noche]

Finally, since all the sequences have less than 6 SGs and, therefore, no more pauses can be assigned, the location of pauses is displayed, as in (7), where pauses are marked with $:

(7) El péndulo había comenzado entonces su oscilación $ y la quietud que reinaba entre nosotros $ era absoluta en el silencio de la noche.

5. Evaluation

In order to assess the performance of ProPause, a comparison between its suggested segmentation and natural speech was made. The later was produced by a speaker, who read aloud a text, including syntactically and prosodically varied sentences, composed of 4979 words. The reading was transcribed with respect to pauses by two phonetically-trained subjects relying on auditory criteria (perceptible silences) and acoustic ones (temporal gaps in waveforms obtained with the speech analysis software Waves+), and compared with the output of the system for the same text.

The deviation of pauses assigned by the system and those made by the speaker was appraised. Data referred to punctuation signs have been excluded of the computation, because we are mainly concerned with the location of orthographically unmarked pauses. Despite being aware of the simplification we made, for the sake of the comparison, it was assumed that all the punctuation signs are associated to a major boundary. In total, 4401 possible pause locations were compared: 4979 word boundaries minus 578 orthographically marked pauses.

Results in Table II show the degree of agreement between the human speaker and ProPause, with respect to phrasing: the table presents the number of pauses made both by ProPause and the human speaker, the number of pauses made by the human speaker but not by ProPause, and, inversely, the number of pauses realized by the system but not by the reader.

We can notice, first, that the number of pauses found in the reading is higher than the number made by the system: 161 in contrast with 129, that is, an increase of 24%. This can be explained, however, by the fact that in some sentences, in which length constraints are not met, the appearance of a pause strongly depends on the speaker's choice.

		Human speaker		
	Pauses	Present	Absent	Total
ProPause	Present	56	73	129
	Absent	105	—	
	Total	161		

Table II. *Degree of agreement between the human speaker and ProPause with respect to pause location.*

On the other hand, the coincidence in the appearance and location of pauses is low: from a total number of 161 pauses realized by the reader, only 56 appear at the same place in both texts, obtaining a 43% of agreement. These results, however, are of limited value since a difference between the speaker and the algorithm does not necessarily imply a mistake on the part of the later, as in (8).

(8) a. Todo cuanto me rodeaba $ parecía haberse transformado mientras me levantaba con la manecilla de oro entre mis dedos.

b. Todo cuanto me rodeaba parecía haberse transformado $ mientras me levantaba con la manecilla de oro entre mis dedos.

'Everything around me seemed to have been transformed while I stood with the little golden key between my fingers'

Both (8a), paused by the speaker, and (8b), paused by the system, are equally acceptable. In order to verify if it is optionality what causes divergences, all cases of discrepancies between the output of the system and naturally produced prosody have been revised by another reader who has marked them as reasonable or impossible, relying on prosodic and syntactic criteria. A pause is considered to be reasonable if it does not violate either prosodic or syntactic requirements, and therefore, it is accepted by the reader's competence.

Table III gives the results grouped in three categories: a) agreement: coincidence in pause assignment between the speaker and the algorithm, b) equivalent phrasing: difference on segmentation yielding equivalent phrasings, and c) non-equivalent phrasing: difference on segmentation resulting on a wrong decision on the part of the system. In addition, results are regrouped in two other categories, involved in the distinction between reasonable and impossible pauses: when phrasing coincides or is equivalent, we are dealing with reasonable pauses, that is, pauses that are acceptable to a listener, in contrast with non-equivalent phrasing, which is interpreted as impossible. It can be noted that from this point of view, the performance of the system increases: the algorithm matches only 43% of the prosodic boundaries made by a speaker but predicts 81% of reasonable pauses.

agreement	equivalent	non-equivalent
56	49	24
reasonable pauses		impossible pauses
105		24

Table III. *Categories found in the comparison between the output of the system and natural prosody.*

If we compare the performance of ProPause with other prosodic segmentation algorithms, despite differences in methodology used to obtain the algorithms and in evaluation procedures, it can be said that its degree of accuracy is correct, although slightly lower than what is obtained in other works. Leaving aside the systems described in Castejón *et al.* (1994) and López (1993), where evaluation results are not offered, in Hirschberg and Prieto (1996) 94.2% correct predictions of phrase boundaries for Mexican Spanish are achieved by means of learning procedures. However, they include in their computation both orthographically marked and unmarked pauses, whereas we are only concerned with the later ones. Related to this, if we take into account orthographically marked pauses in our evaluation, the percentage of correct prediction increases to 92%. Furthermore, in contrast with our approach where syntactic and prosodic factors are included, variables considered by Hirchsberg and Prieto (1996) are mainly prosodic dependent, some of which are incorporated in the CSG unit, for instance, the presence/absence of stress.

Concerning other languages, the algorithm for the Dutch text-to-speech system described in Quené and Kager (1992) predicts 65% of the human prosodic boundaries correctly, but the perfomance improves if different but equivalent phrasings are discarded: 86% of the naturally produced boundaries is matched. This is also in line with results presented in Bachenko and Fitzpatrick (1990) for English: once it is assumed that a difference between a primary and secondary phrase boundary is minimal, the system matched 80% of the boundaries, a result which is very similar to ours: 81%. But in contrast with this study, where overgeneration of pauses is found, ProPause proposes only mandatory pauses and a subset of optional ones.

To conclude, decisions made by ProPause always respect length and syntactic requirements and it is mainly in the domain of optionality where divergences between a phrasing made by a speaker and a phrasing made by the algorithm are found.

6. Discussion and conclusions

Questions addressed to solve the problem of determining the appearance and location of orthographically unmarked pauses in Spanish texts have highligthed some trends. Firstly, there seems to be a minimal and an upward length constraint in reading, which we have modeled in terms of the number of stress groups in a sentence. Prosodic factors affecting the length of IGs have already been stated in other studies on temporal variables in speech (Nespor and Vogel, 1983; Dechert and Raupach, 1980) but in general models use syllables as units. For instance, results on Spanish read texts by Navarro-Tomás (1966) showed a clear preference towards IGs having between 5 and 10 syllables (68%).

Secondly, there are correlations of pause assignment with syntactic units, that we have partially formalized by means of the CSG. Similar to the work of Steedman (1991), we assume that syntactic dependency is a factor which intervenes in the prosodic structure of sentences. Without supporting an isomorphism of intonational structure and syntactic structure, as he does, we agree with the idea of including syntactic and even semantic distinctions in the model. At this stage of the work, however, the CSG is a unit which resembles more 'chunks' as defined by Abney (1991). A chunk is a single content word surrounded by a constellation of function words, matching a fixed template. This author states at least two kinds of relationships between chunks: first, cooccurrence of chunks is determined not just by their syntactic categories, but is affected by the words that head them; and second, the order in which chunks occur is much more flexible than the order of words within chunks. These two restrictions are properly modeled as well by means of the CSG. Thus, we share with Abney the idea that the correspondence between prosodic and syntactic levels can be formalized by means of a unit. The existence of such units is supported by psychological studies such as those performed by Gee and Grosjean (1983).

According to the results, a classification of pauses can be proposed. There are different kinds of pauses: mandatory, optional, impossible and reasonable; and differences between them can be explained by prosodic and syntactic factors. To start with, since there is a length according to which a sentence has to be segmented, it can be said that a pause is mandatory due to prosodic factors. Instead, optionality has to be explained by both prosodic and syntactic factors: there is a length according to which a sentence with a specific number of SGs does not need to be segmented, but if some syntactic requirements are fulfilled, the pause can appear. On the contrary, a pause is impossible when length constraints or categorial cooccurrence restrictions are met —there is a minimal length according to which a sentence cannot be segmented, and there are some CSG boundaries where pauses are not allowed— while a pause is reasonable if it does not violate either prosodic or syntactic requirements. Regarding the implementation of a pause assignment model in an automatic system, we have opted for a rule-oriented approach, in which the selection of units plays an important role. For instance, some of the variables found to be relevant by Hirschberg and Prieto (1996) using the CART techniques, such as the the presence/absence of stress, are incorporated in the system described here from the beginning because of the syntactico-prosodic features of the CSG. Moreover, by using CSG it is possible to predict not only pauses, but also melodic movements. In Aguilar *et al*. (2000) a model of F0 labeling that relates in a direct way melodic movements and linguistic units is proposed. This advantage has been exploited in a text-to-speech system for Galician, described in Fernández-Salgado and Rodríguez-Banga (2000).

To conclude, despite the shortcomings of the work, such as the simplification concerning the relation between punctuation signs and pauses, or the exclusion of the speaking rate as prosodic variable, results concerning the performance of the system suggest that an adequate treatment of text pause assignment is provided.

References

ABNEY, S.A. (1991), "Parsing by chunks" in *Principle-Based Parsing: Computation and Psycholinguistics,* ed. by R. C. Berwick, S. Abney and C. Tenny (Kluwer Ac. Publ., Amsterdam), pp. 257-278.

AGUILAR, L., CASACUBERTA D. and MARÍN R. (2000), "Labeling Melodic Movements at the Stress Group Level", *Catalan Working Papers in Linguistics*, 8, pp. 7-21.

BACHENKO J. and FITZPATRICK E. (1990), "A computational grammar of discourse-neutral prosodic phrasing in English", *Computational Linguistics*, Vol. 16(3), pp. 155-170.

BECKMAN M. and PIERREHUMBERT J. (1986), "Intonational structure in Japanese and English", *Phonology Yearbook*, Vol. 3, 15-70.

CANELLADA M.J. and MADSEN J.K. (1987), *Pronunciación del español*, (Madrid, Castalia).

CASACUBERTA D., MARÍN R. and AGUILAR L. (1998), "Parsing Unrestricted Text into Prosodic Units. A Formal Description", in *Mathematical and Computational Analysis of Natural Language*, ed. by C. Martín Vide, (John Benjamins Publish. Co., Amsterdam), pp. 281-294.

CASTEJÓN, F. ESCALADA, G. MONZÓN, L. RODRÍGUEZ M.A. and SANZ P. (1994), "Un conversor texto-voz para el español", *Comunicaciones de Telefónica I+D*, Vol. 5 (2), pp. 114-131.

CRUTTENDEN, A. (1986), *Intonation*, (Cambridge Univ. Press, Cambridge).

DECHERT H.W. and RAUPACH, M. ed. (1980), *Temporal Variables in Speech*, (Mouton, The Hague).

EMERARD, F. MORTAMET L. and COZANNET A. (1992), "Prosodic processing in a text-to-speech synthesis system using a database and learnig procedures", in *Talking Machines: Theories, Models and Designs*, ed. by G. Bailly and C. Benoit (Elsevier Science Publishers, Amsterdam), pp. 225-254.

FERNÁNDEZ-SALGADO X. and RODRÍGUEZ-BANGA E. (2000), “Proposición de un marco adecuado para el estudio de contornos de F0 para síntesis de voz”, in *Procesamiento del Lenguaje Natural*, 24: 175-181.

FITZPATRICK F. and BACHENKO J. (1989), "Parsing for Prosody: What a Text-to-Speech System Needs from Syntax", *Proceedings of the Annual Artificial Intelligence Systems in Government Conference* (IEEEC Society Press, Washigton), pp. 188-194.

FRENKENBERGER, S. SCHNABEL, B. ALISSALI M., KOMMENDA M. (1994), "Prosodic parsing based on parsing of minimal syntactic structures", *Proc. 2ond ESCA/IEEE Workshop on Speech Synthesis* (New Paltz), pp. 143-146.

GEE J.P. and GROSJEAN F. (1983), "Performance structures: a Psycholinguistic and Linguistic Appraisal", *Cognitive Psychology*, Vol. 15, pp. 411-458.

GILI-FIVELA B. and QUAZZA S. (1997), "Text-to-prosody parsing in an Italian synthesizer. Recent improvements", *Proc. 5th European Conference on Speech Communication and Technology (Rhodes)*, Vol. 2, pp. 987-990.

HIRSCHBERG J. (1991), "Using text analysis to predict intonational boundaries", *Proc. 2nd European Conference on Speech Communication and Technology,* (Genova).

HIRSCHBERG J. and PRIETO P. (1996), "Training intonational phrasing rules automatically for English and Spanish text-to-speech", *Speech Communication*, Vol. 18 (3), pp. 283-290.

LADD R.D. (1987), "A model of intonational phonology for use in speech synthesis by rule", *Proc. of the European Conference of Speech Technology* (Edinburgh), Vol. 1, pp. 21-24.

LADD D.R. (1996) *Intonational phonology*, (Cambridge University Press, Cambridge).

LÓPEZ E. (1993), *Estudio de técnicas de procesado lingüístico y acústico para sistemas de conversión texto-voz en español basados en concatenación de unidades* (Universidad Politécnica de Madrid, Madrid), PhD thesis.

MONAGHAN A.I.C. (1992). "Heuristic strategies for higher level analysis of unrestricted text", in *Talking Machines: Theories, Models and Designs*, ed. by G. Bailly and C. Benoit (Elsevier Science Publishers, Amsterdam), pp. 143-162.

NAVARRO TOMAS T. (1925), "Palabras sin acento", *Revista de Filología Española,* Vol. 12(4), pp. 335-375.

NAVARRO TOMÁS T. (1966), *Estudios de fonología española*, (Casa de las Américas, Nueva York).

NESPOR M. and VOGEL I. (1983), "Prosodic Structure above the Word", in *Prosody. Models and Measurements*, ed. by A. Cutler and R. D. Ladd (Springer Verlag, Heidelberg), pp. 123-140.

NESPOR M. and VOGEL I. (1986), *Prosodic phonology*, Foris Publications, Dordrech. Trad.: *La prosodia*, Visor, Madrid, 1994.

NOOTEBOOM, S.G.. BROKX J.P.L and ROOIJ J.J.de (1978), "Contributions of prosody to speech perception", in *Studies in the Perception of Language*, ed. by Levelt W.J.M. and Flores d'Arcais G.G. (John Wiley, Chichester), pp. 75-107.

O'SHAUGHNESSY D. (1990), "Relations between syntax and prosody for speech synthesis", *Proc. of the ESCA Workshop on Speech Synthesis*, (Autrans).

PIERREHUMBERT, P. (1980), *The Phonology and Phonetics of English Intonation*, (MIT, Cambridge), PhD thesis.

QUENÉ H. and KAGER R. (1992), "The derivation of prosody for text-to-speech from prosodic sentence structure", *Computer Speech and Language*, Vol. 6, pp. 77-98.

QUILIS A. (1993), *Tratado de fonética y fonología españolas*, (Gredos, Madrid).

SOSA J.M. (1999) *La entonación del español*, Cátedra, Barcelona.

STEEDMAN M. (1991), "Structure and Intonation", *Language,* Vol. 67(2), pp. 260-296.

VANNIER, G. LACHERET-DUJOUR A., VERGNE J. (1999) “Pauses location and duration calculated with syntactic dependencies and textual considerations for t.t.s. system”. In *Proceedings of the 14th International Congress of Phonetic sciences (ICPhSC’99*), San Francisco, 1999.

VERGNE J. (1993), "Syntax as clipping blocks: structures, algorithms and rules", *Boletín SEPLN*, Vol. 13, pp. 179-197.

Acknowledgements

The experiments in section 3 are based on and reinterpreted from the paper presented by the authors in the *XIIth Conference on Natural and Formal Languages,* held in La Seu D'Urgell in September 1996. A formal treatment of the pause assignment model was presented at the *II International Conference on Mathematical Linguistics*, in Tarragona in March 1996; and the evaluation was presented at *Eurospeech'97,* in Rhodes (Greece) in September 1997. Thanks to all the audiences for their comments.