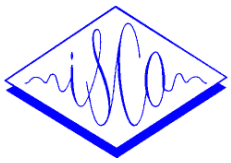

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

ANALYSIS OF ACCENT AND INTONATION IN SPANISH BASED ON A QUANTITATIVE MODEL

*Hiroya Fujisaki**, *Sumio Ohno**, *Kei-ichi Nakamura**, *Miguelina Guirao*** and *Jorge Gurlekian***

* Department of Applied Electronics, Science University of Tokyo
2641 Yamazaki, Noda, 278 Japan
** Laboratorio de Investigaciónes Sensoriales, CONICET
Marcelo T. de Alvear 2202 - 4o Piso, Buenos Aires, Argentina

ABSTRACT

The aim of the present study is to obtain an analytic and quantitative description of the prosodic characteristics of spoken Spanish and to find out the relationship between the description and the underlying linguistic and paralinguistic information. This paper presents results of our preliminary effort to apply a method of analysis of F_0 contours that has been proposed and proved to be valid for Japanese and several other languages. Analysis of utterances of declarative sentences of Spanish indicates the basic validity of the method and provides useful findings on the prosodic characteristics of accent and intonation in Spanish.

1. INTRODUCTION

The accent (*acento*) in Spanish has been commonly considered as the stress accent and has been qualitatively discussed within the scope of a word. On the other hand, the intonation (*entonación*) has been considered as changes in pitch related to differences in the type and function of a sentence such as statement and question [1], but its features have not been analyzed in quantitative terms.

Our study is aimed at obtaining an analytic and quantitative description of the prosodic characteristics of spoken Spanish and thus finding the relationship between the description and the underlying linguistic and paralinguistic information [2]. In particular, we try to explore the possibility of applying a method of analysis of F_0 contours which has originally been proposed and proved to be valid for Japanese and extended to several other languages [3]. This paper describes the results of our preliminary analysis conducted on declarative sentences with relatively simple syntactic structures uttered by two native speakers [4].

2. SPEECH MATERIAL AND METHOD OF ANALYSIS

As the initial step in exploring the validity of applying or extending our method of analysis, we tried to apply it to a set of declarative sentences of various length, first with a neutral statement intonation and then with several different patterns of intonation expressing paralinguistic modification of the original statement such as declaration, interrogation, suspicion, command and admiration.

The speech material for each speaker consists of the following three sets of utterances. Set A consists of one utterance each of 10 different sentences (Utterance (1) to (10)) read with a neutral intonation of statement. Starting with four word sentences, longer utterances are produced by adding new words up to a total length of 21 words. These utterances are used to study the basic characteristics of the F_0 contour for the neutral statement intonation and the effect of utterance length on it. Set B consists of six different utterances of the same sentence (a 12-word sentence taken from Set A) produced with an emphasis on one of six different positions. These utterances are used to study the effect of emphasis on the F_0 contour. Finally, set C consists of five utterances of the same sentence (same as that for Utterance (9) of set A) produced with five different types of intonation other than the neutral one, corresponding respectively to the expression of declaration, interrogation, suspicion, command, and admiration. Table 1 lists all the utterances used in the present study. In order to be able to analyze F_0 contours as uninterrupted curves, these utterances contain only vowels and voiced consonants. An underline in each of the utterances of Set B indicates the word or words to be emphasized.

These utterances were produced and recorded by two native speakers (one male and one female) of Castillian Spanish from Buenos Aires, who read the list six times at a normal speech rate of approximately six syllables per second without pauses in each utterance.

The recorded material was digitized at 10 kHz with 12 bit precision for further analysis. Fundamental frequencies were extracted by a modified autocorrelation analysis of the LPC prediction residual. The F_0 con-

Table 1. List of Spanish utterances analyzed.

Set A
(1) La abuela le da.
(2) La abuela le da un melon.
(3) La abuela de Lola le da un melon.
(4) La buena abuela de Lola le da un melon a la nena.
(5) La buena abuela de Lola le da un melon y una banana a la linda nena.
(6) La buena abuela de Lola le da un melon y una banana a la linda nena en el dia de hoy.
(7) Le da un melon.
(8) Le da un melon y una banana.
(9) Le da un melon y una banana a la nena.
(10) Le da un melon y una banana a la linda nena.

Set B
(11) La buena abuela de Lola le da un melon a la nena.
(12) La buena abuela de Lola le da un melon a la nena.
(13) La buena abuela de Lola le da un melon a la nena.
(14) La buena abuela de Lola le da un melon a la nena.
(15) La buena abuela de Lola le da un melon a la nena.
(16) La buena abuela de Lola le da un melon a la nena.

Set C
(17) Le da un melon y una banana a la nena. (declaration)
(18) Le da un melon y una banana a la nena. (interrogation)
(19) Le da un melon y una banana a la nena. (suspicion)
(20) Le da un melon y una banana a la nena. (command)
(21) Le da un melon y una banana a la nena. (admiration)

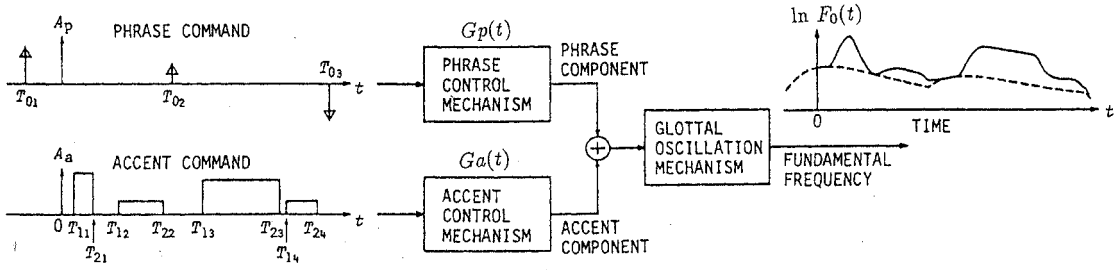

Fig. 1. A quantitative model for the process of F_0 contour generation.

tours were further analyzed by the method of Analysis-by-Synthesis using a quantitative model for the process of F_0 contour generation [1].

As shown in Fig. 1, the model is based on the assumption that the F_0 contour of a declarative sentence in the logarithmic frequency scale can be considered as the sum of two kinds of components, i.e., the phrase components and the accent components, and that the phrase component is the response of a critically-damped second-order linear system to an impulse-like phrase command, while the accent component is the response of another critically-damped second-order linear system to a step-like accent command. Thus an F_0 contour of a spoken sentence is represented by the following equation:

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j \{Ga(t - T_{1j}) - Ga(t - T_{2j})\}, \quad (1)$$

$$Gp(t) \begin{cases} = \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ = 0, & \text{for } t < 0, \end{cases} \quad (2)$$

$$Ga(t) \begin{cases} = \min[1 - (1 + \beta t)\exp(-\beta t),\ \gamma], & \text{for } t \geq 0, \\ = 0, & \text{for } t < 0, \end{cases} \quad (3)$$

where $Gp(t)$ represents the impulse response function of the phrase control mechanism and $Ga(t)$ represents the step response function of the accent control mechanism. The symbols in these equations indicate

- Fb : baseline value of an F_0 contour,
- I : number of phrase commands,
- J : number of accent commands,
- Ap_i : magnitude of the i-th phrase command,
- Aa_j : amplitude of the j-th accent command,
- T_{0i} : timing of the i-th phrase command,
- T_{1j} : onset of the j-th accent command,
- T_{2j} : end of the j-th accent command,
- α : natural angular frequency of the phrase control mechanism to the phrase command,
- β : natural angular frequency of the accent control mechanism to the accent command,
- γ : a parameter to indicate the ceiling level of the accent component (generally set equal to 0.9).

Parameters α and β characterize dynamic properties of the respective laryngeal mechanisms for phrase and accent control, and are considered to remain fairly constant within an utterance. They may not change widely across different utterances by one speaker, nor across different speakers. The parameter Fb represents the approximate lower limit of the voice register for a particular utterance and thus do not vary appreciably across utterances of the same speaker but may vary from speaker to speaker.

3. RESULTS OF ANALYSIS

3.1 Utterances with Neutral Intonation

Figure 2 shows one example each of the F_0 contour analysis of the six utterances of Set A by the male speaker JG. Each panel shows, from top to bottom, the speech waveform, the observed F_0 contour as a sequence of + symbols, the best approximation generated by the model as a curve in a solid line, the estimated phrase components as a curve in a dashed line, the estimated baseline value Fb in a dotted horizontal line, and the estimated accent commands. The difference between the solid line and the dashed line corresponds to the estimated accent components.

The six examples in Fig. 2 illustrate the very close approximations that can be achieved by the model. The number of phrase components is one for Utterances (1) and (2), two for Utterances (3) and (4), four for Utterance (5), and five for Utterance (6). For instance, Utterance (6) is seen to have phrase commands corresponding to: *La buena abuela de Lola / le da un melon / y una banana / a la linda nena / en el dia de hoy.*

The utterance-initial phrase command leads the segmental onset of the utterance by 130 $\sim$ 220 msec, and its magnitude is approximately 0.5 regardless of the length of the utterance. An utterance-medial phrase command leads the segmental onset of the corresponding prosodic phrase by a similar amount of time or less, and its magnitude is always less than that of the utterance-initial phrase command, being about 0.3 for the second phrase command, and is generally smaller for the third.

The accent command does not occur at every word of the utterance, but only at certain words or phrases. For instance, in Utterance (6) it occurs on all the adjectives, nouns and verbs but not on particles and prepositions. As in the case of Japanese, we may define a prosodic word that corresponds to each of the accent command. In this definition, particles and prepositions are to be combined with the following word to form prosodic words such as *la buena, de Lola, le da, un melon, una banana, a la linda, en el dia de hoy.* The amplitude of the accent command in the examples of Fig. 2 takes a value within the range from 0.12 to 0.55. Within an utterance, a speaker tends to put an emphasis on certain prosodic words by default, so that the amplitudes of their accent commands are considerably higher than others in the same utterance. We shall refer to the former as 'prominent' and the latter as 'non-prominent' prosodic words, respectively. Figure 3 shows the distributions of the accent command amplitudes of these two groups of prosodic words. Since the level of accentuation varies from utterance to utterance, and also varies even within an utterance, the two distributions show a certain amount of overlap, but their distinction is clear within an utterance because of contrast.

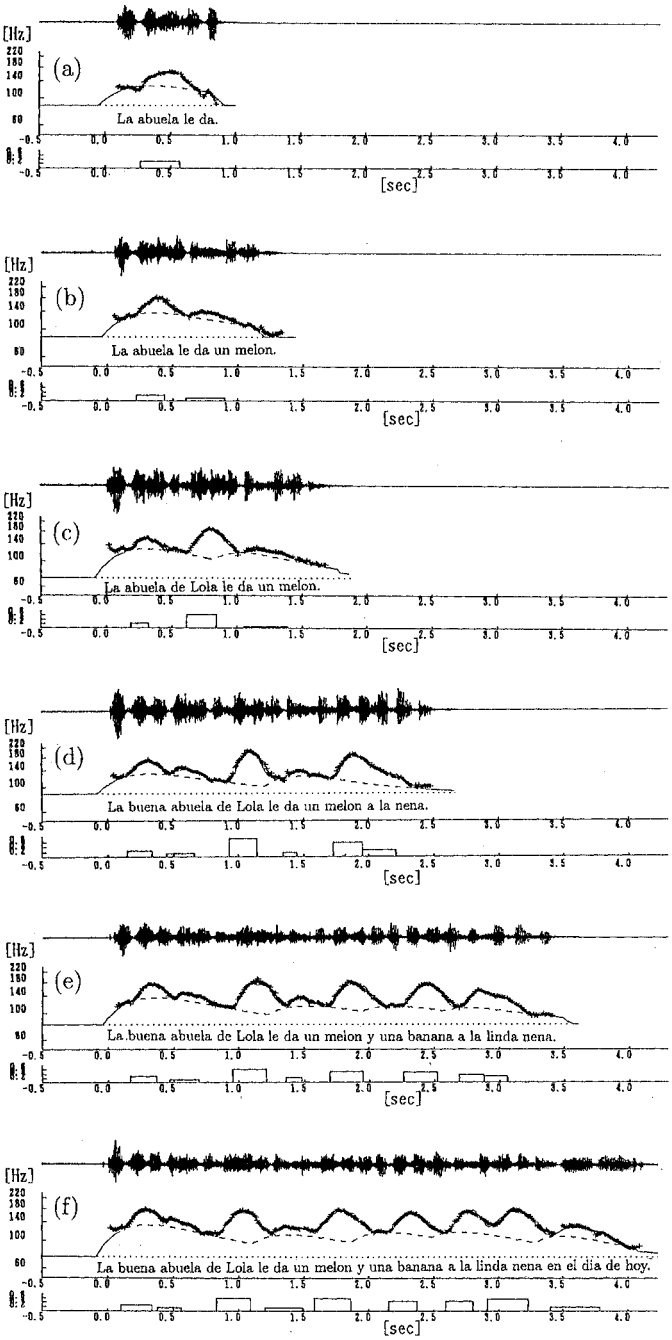

Fig. 2. Examples of F_0 contour analysis of Spanish utterances with neutral intonation (Set A).

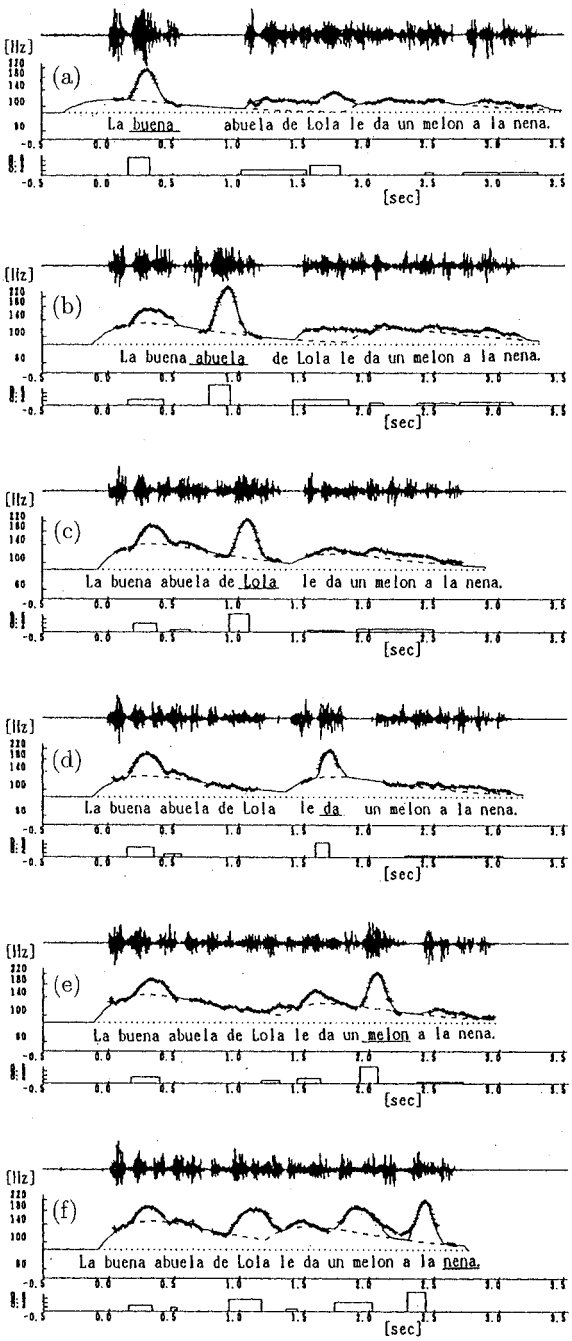

Fig. 4. Examples of F_0 contour analysis of Spanish utterances with specified emphasis (Set B).

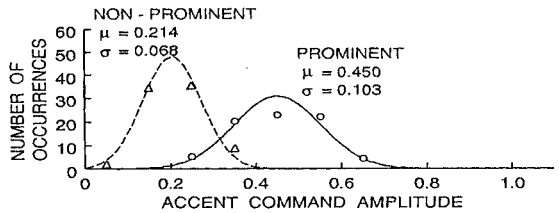

Fig. 3. Distributions of accent command amplitude for prominent and non-prominent prosodic words in Spanish utterances with neutral intonation.

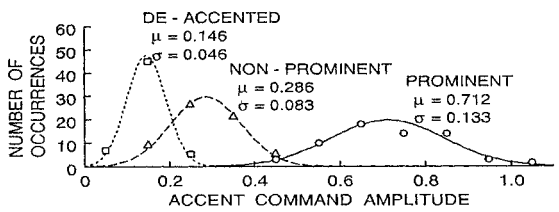

Fig. 5. Distributions of accent command amplitude for prominent / non-prominent / de-accented prosodic words in Spanish utterances with specified emphasis.

3.2 Utterances with Intended Emphasis

The six panels (a) ~ (f) in Fig. 4 show one example each of the F_0 contour analysis of the six utterances of Set B by the same speaker as in Fig 2. Comparison of Fig. 4 with Fig. 2 (d) clearly indicates that the accent command amplitude for the prosodic words with intended emphasis is considerably higher than that of the default emphasis in the utterance of Fig. 2 (d). In addition, the accent command amplitude on other prosodic words are markedly reduced by contrast in most cases except for the utterance-initial adjective '*buena*' which receives a similar degree of emphasis to that in a neutral intonation. The utterance in panel (f) is an exception in that, in addition to the high amplitude of the accent command for the intended emphasis on the prosodic word '*a la nena,*' the accent command amplitude is rather high at two other positions, though not quite as high as that for the intended emphasis. From the point of view of the accent command amplitude, therefore, prosodic words in utterances with intended emphasis fall broadly into three groups: prominent, non-prominent, and de-accented. Figure 5 shows the distributions of the accent command amplitudes of these three groups of prosodic words. While there is very little overlap between the distributions of prominent and non-prominent prosodic words, certain amount of overlap is found between those of non-prominent and de-accented prosodic words. Comparison of Fig. 5 with Fig. 3 indicates that in utterances with specified emphasis, the accent command amplitude is higher not only in the prominent group but also in the non-prominent group of prosodic words.

Analysis of variance shows that the difference in the mean values of accent command amplitude for the same word in Set A utterances (neutral intonation) and Set B utterances (with intended emphasis) are highly significant at all positions in utterances of speaker JG, and at all positions except at the utterance-initial '*buena*' and utterance-medial '*Lola*' in speaker MG, who tends to put an emphasis on these words by default.

3.3 Utterances Expressing Various Intentions

F_0 contour analysis has also been conducted on utterances of Set C. The literary meaning of the sentence is "give a melon and a banana to the baby (girl)" without a subject. Because of the absence of subject, various meanings can be added by intonation as in the following utterances:

(17) I give a melon and a banana to the baby. (a declarative intonation)

(18) Do you give a melon and a banana to the baby? (with an interrogative intonation)

(19) Do you really give a melon and a banana to the baby? (with a suspicious intonation)

(20) Give a melon and a banana to the baby! (with an imperative intonation)

(21) Oh, you give a melon and a banana to the baby! (with an admiring intonation)

As far as the current data is concerned, the characteristics of the declarative intonation is quite similar to those of the neutral intonation. We may therefore use Utterance (17) as a reference for comparison. The characteristics of other intonation types of utterances by Speaker JG can be briefly summarized as in Table 2.

It should be noted, however, that there exist considerable individual differences in the F_0 contour characteristics especially in intonation types expressing command and admiration. A more extensive study is apparently necessary to find out the essential characteristics of each type of intonation and their differences.

Table 2. Characteristics of various types of intonation.

Intonation	Characteristics
Interrogation	Slower speech rate, reduced accent command amplitudes except the one on the utterance-final word.
Suspicion	Normal speech rate, increased accent command amplitudes for all the prosodic word *nena* falls on the last syllable and is especially large.
Command	Faster speech rate, higher F_0 and generally increased accent command amplitudes, a large and gradual drop in F_0 toward the final word of the utterance.
Admiration	Very low speech rate, higher F_0, a phrase command for each prosodic word, increased accent command amplitude, especially on the initial syllable of the final word *nena.*

4. SUMMARY AND CONCLUSION

We have described our preliminary effort toward the analysis of prosodic characteristics of spoken Spanish. Our results demonstrate the validity of the method of F_0 contour analysis originally proposed and applied to Japanese and several other languages. Analysis of F_0 contours of Spanish utterances with various positions of emphasis and various types of intonation has yielded some quantitative findings on the prosodic characteristics of accent and intonation in Spanish.

REFERENCES

[1] Navarro, T. T., *Manual de Pronumciación Española*, Centro de Estudios Históricos, Madrid (1926).

[2] Fujisaki, H., "From information to intonation," *Proceedings of the 1993 International Symposium on Spoken Dialogue, Tokyo*, pp. 7–18 (1993).

[3] Fujisaki, H. and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *J. Acoust. Soc. Jpn (E)*, Vol. 5, No. 4, pp. 233–242 (1984).

[4] Fujisaki, H., K. Nakamura, M. Guirao and J. Gurlekian, "Computational modeling of accent and intonation in declarative sentences of Spanish," *J. Acoust. Soc. Am.*, Vol. 95, No. 5, Pt. 2, pp. 2949 (1994).