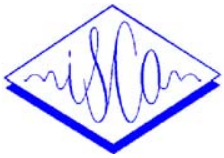


6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A SEMI AUTOMATIC METHOD FOR THE CHARACTERIZATION OF SPANISH INTONATION CONTOURS

Jorge A. Gurlekian**,** Laboratory of Sensory Research–CONICET–UBA, jag@lis.edu.ar, www.lis.secyt.gov.ar
Marcela Leticia Riccillo, Facultad de Ciencias Exactas y Naturales, UBA. gamma@ciudad.com.ar
Alejandro Renato**,** Facultad de Medicina, UBA, arenato@dc.uba.ar
Jose Alvarez, Facultad de Ciencias Exactas y Naturales, UBA. jalvarez@dc.uba.ar

10.21437/Eurospeech.1999-187

ABSTRACT

The purpose of this work is to present a method to characterize Spanish Intonation Patterns (SIP) and to describe its principal features. A semi automatic method was developed, which consisted in the stylization of SIP and coding of pitch movements. A preliminary corpus of 120 sentences were collected and recorded by two native speakers of Buenos Aires (one female, one male). In the stylization process, smoothing of the row F0 contours in two stages and a polynomial approximation by straight-line segments were applied. A perceptual test assessed the intonation quality obtained for sentences using the stylized f0 contours. Subjects reported no differences with natural ones. In the coding process, the micro and macro prosodic features of the SIP were represented. Spanish ToBI notations were defined and combined with a hierarchy structure according to the ERB-rate scale. This notation strategy was applied to the corpus, resulting in a set of sentence labels. We applied this set of sentence labels to an equivalent text corpus and tested the quality of the predicted F0 contours by concatenating interpolated demi-syllables. The quality of the synthetic speech obtained, suggest the convenience to extend the method to a larger corpus.

INTRODUCTION

Current models for intonation look at the fundamental frequency contours either as:

a) an interaction of contours at different levels such as the accent and phrase contours. The command-filter model of Fujisaki, (1), or
b) a series of target values interpolated by a transition function, Pierhumbert and Hirst, (2,3) or a series of rising and falling instructions, 't Hart (4).

One of the initial descriptions on Spanish intonation was made by Navarro Tomás, (5). It determines that the intonation has a melodic unity which are the lines of tones that go with the phonic group, (the syllable chain with one or more than one strong accent, where the links are the initial silent and the juncture or two junctures). Intonation variations were described according to modality and its different realizations.
Other relevant study was made by Sosa (6) who implemented the auto-segmental intonation model for Spanish.

Based on the actual intonation models we developed a semi automatic method to describe Spanish intonation, organized in the following steps: corpus development, digitalization and analysis, stylization and codification.. We choose the ToBI notation system combined with the ERB-rate scale to produce a richest representation of the micro and macro prosodic features of the F0 contours. The stylization was intended to facilitate the labeling of prosodic markers.

METHOD DESCRIPTION

Corpus development. A base sentence was systematically varied to produce a corpus of 120 Spanish utterances representing the prototypical variations in intonation. Two speakers -one female and one male- both natives of Buenos Aires city were recorded in three sessions. Intonation variations according to Navarro (5) were applied to a base sentence /el aBuelo le Da un elaDo de limon al nene/, "grandfather gives a lemon ice-cream to the boy", which has all voiced sounds to avoid discontinuities. Nevertheless during its development we found that the method could be extended to unvoiced segments. The intonation variations were the following:
Length. Five different lengths were applied to the base sentence to analyze its influence.
Modality. The speakers produced the base sentence uttered as affirmative, interrogative, exhortative, negative, and exclamative sentences.
Partial enumeration. The contour shape of the base sentence was examined according to his position within a longer sentence. For this purpose two carrying sentences "El sol ilumina el parque" and "Los chicos juegan alegremente" were defined and then three new sentences were prepared focusing the basic sentence at the beginning, middle and final position relative to the two carrying sentences.
Complete enumeration. The influence of coma in the contour shape was explored. The clause sentence "mi buen amigo" was intercalated at the middle and final of the base sentence.
Topic. Four cases were defined. The speakers answered with the base sentence to questions related to segments contained inside the sentence. (eg. Who is giving the lemon ice-cream to the boy?).

Digitalization and Analysis. 16 bits and 10kHz S/R for digitalization of the speech waveform were employed. The speech analysis system of the Laboratory of Sensory Research allowed the Cepstral analysis for F0 and RMS estimation every 10 msec to define the temporal contours.

Stylization process. The first step for the stylization was to define the F0 values to zero during the pauses and at the beginning and end of the utterance according to the RMS contour for a given RMS threshold. The automatic stylization consisted in two steps: the first one was the elimination of erroneous F0 values according to a maximum difference allowed with the previous value. The second one was to smooth the contour by a convolution filter (9) to remove high frequencies which were not perceptually relevant. The parameters of the convolution filter were adjusted until the resulted contour was acceptable to a panel of listeners. This method was tested by a group of listeners who compared the stylized versions with the original ones. The synthesis was made by a program based on Klatt's synthesizer (10).

The lineal polynomial approximation. The values obtained in the previous step were interpolated by lineal segments. For this election we consider the results obtained by 't Hart (8) who compared straight vs parabolic interpolation and obtained no significant differences.

Codification of the tonal movements: A labeling method based on ToBI (7) combined with a hierarchical structure according to the ERB rate (equivalent-rectangular-bandwidth rate scale) (8) was used. ToBI resulted the most convenient method because it considers the microprosodic features. In order to improve the comparison of contours with different tones heights, which are necessary to regenerate the contours from the codificated chains, we obtained the local maximum and local minimum values and classified them in levels. To make this normalization we transformed the values in the *ERB-rate* (equivalent-rectangular-bandwidth rate scale), which can be defined as

$$ERB(x) = 16{,}7 \ * \ \log_{10}(1 + (x / 165{,}4))$$

where x: Frequency in Hz and ERB(x): ERB value

By means of this transformation F0 values from 20 to 650 Hz were converted to levels from 1 to 12.

Example The figures shows the base sentence "El abuelo le da un helado de limón al nene." uttered by the male speaker when answering the topic question "¿Qué le da el abuelo al nene?" (what's the grandfather giving to the boy?).

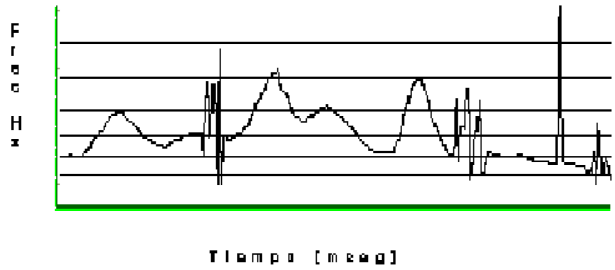


Figure 1. Original contour, with pauses and spurious frequencies.

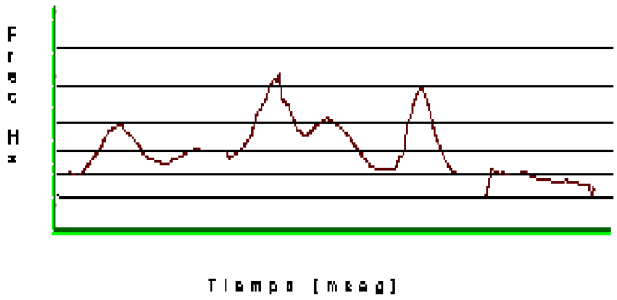


Figure 2. Idem Figure 1. After de first part of the stylization

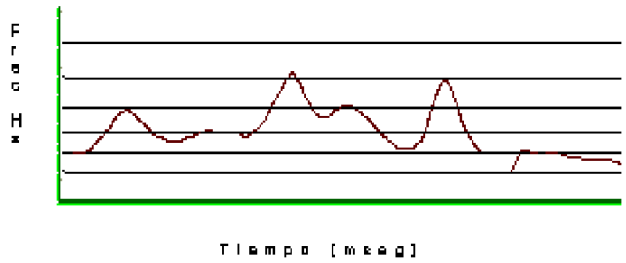


Figure 3 Idem Figure 1. After de second part of the stylization

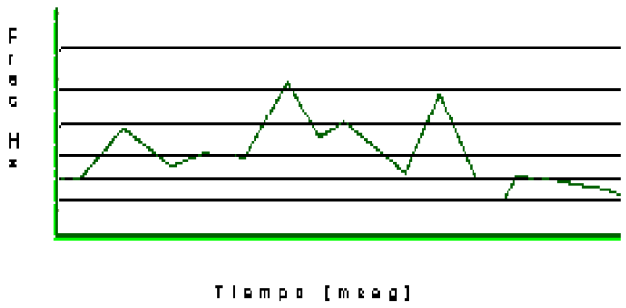
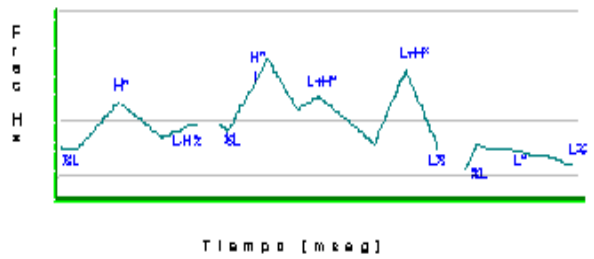


Figure 4. Idem Figure 1. After de lineal approximation



El a **bue** lo le**da** un he **la**do deli**món** al **ne** ne
%5L 6H* 5L-6H% %5L 7H* 6L+7H* 5L+7H* 5L% %4L 5L* 4L%

Figure 5. Idem Figure 1. After de labeling with the stressed syllables and final codification according to ERB's.

EXPERIMENTAL RESULTS

Once the method was implemented we tested it with the intonation contours derived from the corpus mentioned before. En total the *120 utterances came from* 60 female and 60 male emissions from the three sessions for each of the twenty designed sentences derived from the base sentence..

The contours were corrected, stylized and labeled. For the codification we needed to manually recognize the location of the stressed syllables. For this purpose we used the RMS energy contour. The maximums of this function corresponded roughly to the location of the vowels and the minimums to the consonants.

Shapes of the stylized patterns

As a graphic summary , we present some examples of the the main shapes of the intonation patterns of the Spanish spoken in Buenos Aires which were detected from the present corpus.

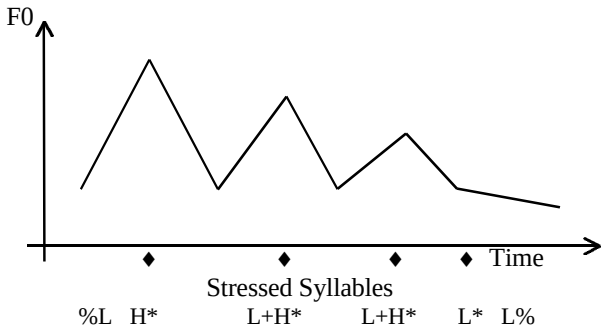


Figure6. Typical pattern for affirmation

This shape is similar in the sentences with other modalities such as negation and exhortation.

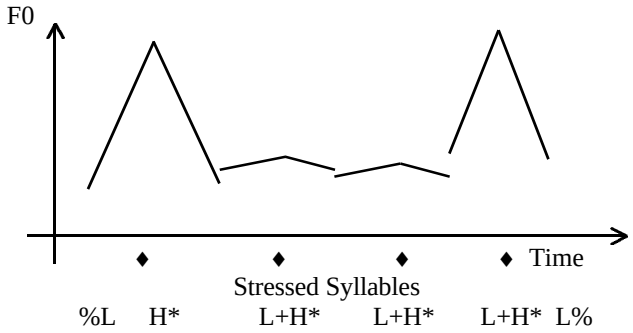


Figure 7. Typical pattern for interrogation

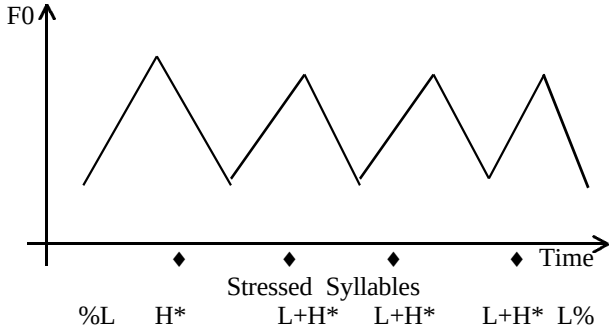


Figure 8. Typical pattern for exclamation.

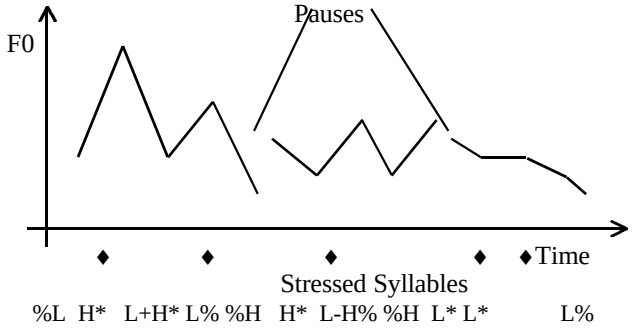


Figure 9. Typical pattern for Partial and complete Enumeration

These characteristics were also found in the category of length variations.

Topic

In the sentences where topic was the target, the prominent maximums are located in the segments that emphasize the subject of the sentence. Some times a brief pause appears before or after these segments that enhance the meaning of the sentence..

CONCLUSIONS

In this work a semi-automatic method for the characterization of Spanish intonation patterns is presented.(SIP). They were obtained by digital signal processing. The sentences were recorded and after digitalization and analysis the fundamental frequency contours were obtained.
The stylization process allowed to parameterize the fundamental frequency contour and control the fundamental frequency during pauses. It was also used in the labeling process to plot the RMS function in order to align the stressed syllables and also to locate the maximum and minimum values in their correspondent levels.

A- The stylization process suppress the prosodic information non relevant from the contours, sustaining the perceptual information identical to the original contours.

B- The stylization process reduced the storage space for the contours. The stylized contours occupy only 15% of the original contours. This eases its representation, reproduction and interpretation.

C- By the perceptual tests we can test the reliability of the stylized contours which are perceptually identical to the originals.

D- The labeling method represents the general characteristics of the contours (macro-prosodic features) and the F0 values related with the position of the stress syllables (micro-prosodic features). This allow a correct comparison between F0 patterns, considering the frequency ranges. For example the patterns that generate the same coding marks with F0 values in different frequency ranges, or between local maximums of different heights in the same contour.

E- After the application of the method to the contours of the corpus we found singular shapes. This allowed us to characterize them and differentiate them according to the intonation variation which they represented.

F- Particular shapes of the intonation contours were obtained which allow to identify typical intonation patterns of the Spanish spoken in Buenos Aires for affirmative sentences with and without pauses, and for interrogatives and exclamatives

G- The observed characteristics in the intonation contours of this corpus after the application of the proposed method were tested with an adaptation of the Klatt formant synthesizer provided by the Laboratory of Sensory Research.

Future directions- Labelling and training: In a future project we plan to extend this method to a larger corpus with different types of variations and made it available for reference testing and comparisons. One possible direction is to automatize the labeling process. Generating speech: We plan to apply this method to predict the shapes of the intonation contours of any other corpus. These results could be used to synthesize sentences with a storage economy, easy management and a principally a better intonation.

REFERENCES

[1] Fujisaki Hiroya, "From Information to Intonation", Lecture at Laboratorio de Investigaciones Sensoriales, on the occasion of its XXV the Anniversary, 1993.

[2] Pierrehumbert Janet, "Synthesizing intonation", J. Acoust. Soc of Am, Vol 70, N 4, 985-995, 1981.

[3] Hirst Daniel, Campione Estelle, Flachaire Emmanuel, Véronis Jean, "Stylisation and Symbolic coding of F0: A Quantitative Model", ESCA Workshop on Intonation: Theory, Models and Applications, Athens Greece, 71-74, 1997.

[4] Hart ('t) Johan, Nico Willems, Collier René, "A synthesis scheme for British English intonation", Journal Acoustical Society of America, Vol 84, N 4, 1250-1262, 1988.

[5] Navarro Tomás T., *"Manual de Entonación Española"*, Ed. Guadarrama, 1974.

[6] Sosa Juan Manuel, “Fonetica y Fonologia de la entonacion del espanol Hispanoamericano”, University of Massachusetts, Department of Spanish and Portuguese, 1991.

[7] Beckman Mary, Ayers Gayle, "Guidelines for ToBI Labeling", obtained from internet, 1994.

[8] Hermes D.J., Van Gestel, "The frequency scale of speech intonation", Journal Acoustical Society of America, Volumen 90 N 1, 97-102, 1991.

[9] Balmer Leslie, *"Signal and Systems. An Introduction"*, Prentice Hall, 1987.

[10] Klatt Dennis, "Software for a cascade/parallel formant synthesizer", Journal Acoustical Society of America, Vol 67, 971-995, 1980.