MODELIZATION OF INTONATION PATTERNS IN SPANISH FOR AUTOMATIC RECOGNITION

Juan María Garrido and Francesc Gudayol

Laboratori de Fonètica, Universitat Autònoma,
Barcelona, Spain

This paper describes a simple method to derive stylized representations from raw F0 contours. The results of a study on F0 contours in Spanish sentences using this method are also presented. Finally, application of obtained models to automatic speech recognition is discussed.

1. INTRODUCTION

Modelization of prosodic information for automatic synthesis and recognition systems can be made in several ways. An usual approach has been the analysis of fundamental frequency (F0), duration and amplitude of syllabic *nuclei* in sentences, in order to detect stressed syllables and clause boundaries. Intonative information can also be extracted from a representation of different F0 levels in the syllables.

Such an approach has demonstrated to be very effective in synthesis systems (as the one developed by Pierrehumbert [3]). But in recognition systems, although important results have been achieved, some questions, as the automatic detection of syllablic nucleus, have not still been entirely solved. Results reported by Vaissière [4] and Mertens [1] are good examples of this fact.

This paper describes a simple method of representation of F0 contours (section 1), and a study of melodic patterns in Spanish sentences using stylized contours obtained with this procedure (section 3). This study has been achieved as a primary step in the development of an automatic pitch contour recognizer for Spanish. The use of the obtained results in automatic recognition is evaluated in section 4.

2. STYLIZATION PROCESS

Process of stylization included various stages:

1) Extraction of F0 values.

2) Obtention of stylized representations, by saving only the values corresponding to time and F0 at beginning, end and inflection points of each sentence. Inflection points can be defined as points in the F0 contour were slope changes its sign (from positive to negative, or *vice versa*). Straight lines traced between these points form the stylized representations for each contour.

3) Finally, simple time and frequency normalizations were performed, in order to avoid variations due to intrinsic F0 and speech rate for different speakers. In frequency normalization, initial F0 values of each contour were set to 0, and values corresponding to the remaining points were made relative to it. Time normalization sets to 0 and 1 time values corresponding to beginning and end of F0 trace respectively; other time values in the representation are made relative to these reference points.

This stylization procedure is very straightforward and can be easily

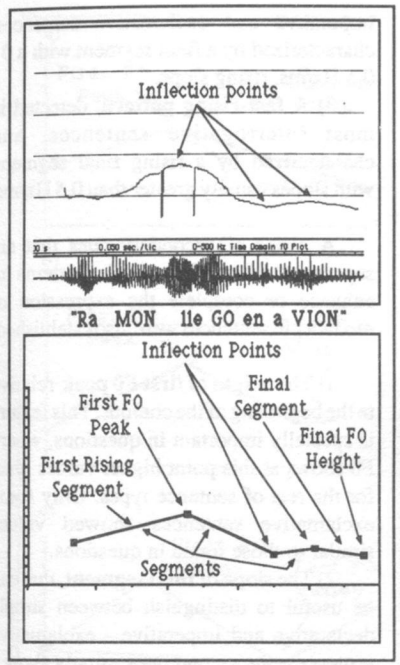

FIGURE 1: An example of original F0 trace, and its correspondant stylized representation

automatized if desired, although it has been applied manually in this study. An example of stylized representation, and its corresponding F0 contour, is shown in Figure 1.

Some remarks can be made about this method:

1) A whole-sentence approach has been preferred to a syllable approach in this analysis procedure. The F0 trace is divided into several **segments**, defined as the straight lines that link two contiguous inflection points. Obtained F0 peaks can correspond or not to stressed syllables.

2) Only F0 variations are analyzed. Duration and amplitude of syllables are considered to be important for transmission of stress information, but they seem secondary cues in intonation analysis.

3) Micromelodic variations are not considered in this method. Segments showing increases or drops in F0 less than 10 Hz between its beginning and its end are not taken account of.

3. EXPERIMENTAL PROCEDURE

The described procedure was applied to stylize a set of F0 contours extracted from simple Spanish sentences. The analysis of the obtained representations carried to the definition of some typical F0 patterns used in Spanish for expression of sentential modality.

3.1. Design and Recording of *Corpus*

A set of 54 sentences was constructed representing six different modality types, adapted from the traditional classification of Navarro Tomás [2]: a) enunciative; b) interrogative type I; c) interrogative type II; d) interrogative type III; e) exclamative; and f) imperative. Each subset included different tokens of sentences with changes in number and position of stressed syllables.

Sentences contained each only one intonative group, and were formed exclusively by voiced sounds, to provide the analysis of complete F0 contours. They were embedded into brief dialogues to facilitate the production of a more natural intonation while reading. Dialogues were read by 4 Spanish speakers, 2 men and 2 women, in a sound isolated booth, and recorded on high-quality audio tape.

3.2. Analysis Procedure

Sentences to be analyzed were then low-pass filtered, digitized at a sample rate of 10 Khz, and stored. F0 contours calculations were performed by means of a pitch detector based on an auto-correlation technique, available in MacSpeech Lab II, a commercial speech analysis software for Macintosh.

The procedure described in section 2 was applied to each F0 contour in order to obtain the stylized representations.

Previous studies [5] showed that initial and final parts of F0 contours contain most intonative information. To take account of this fact, slopes of initial and final segments for each sentence were calculated.

3.3. Statistic Analysis

Two sub-groups of sentences were formed for each modality type, according wether the slope of final segment was rising or falling. A simple statistic analysis was then performed to extract mean values of the slope of first and last sentence segment for each sub-group. Means were also calculated for F0 values at the first peak (that is, at the end of the first rising segment), and at the end of the contours. The results of this analysis are shown in Table A.

3.4. Pattern Classification

According to these results, the obtained F0 stylized contours were classified into three basic patterns (see figure 2):

1) **A falling pattern**, for all declarative and some interrogative, imperative and exclamative sentences, characterized by the presence of a final segment with falling slope.

2) **A slow-rising pattern**, for some imperative and exclamative sentences, characterized by a final segment with a 0 - 0,5 Hz/ms. rising slope.

3) **A fast-rising pattern**, detected in most interrogative sentences, and characterized by a rising final segment, with slopes usually greater than 0,5 Hz/ms.

A series of secondary cues that are superimposed to these basic patterns to enhance or complete the expression of modality in sentences were also established:

1) The **height of first F0 peak** relative to the beginning of the contour. This feature is specially important in questions, where F0 shows at this point higher values than for the rest of sentence types. Only some exclamative sentences showed values similar to those found in questions.

2) The **slope of final segment**, that can be useful to distinguish between simple declarative and imperative - exclamative sentences: the second ones usually show a more abrupt slope (less than -1 Hz/ms.) than the first one.

3) The use of **some special F0 contour forms**, as the "circumflex final", or the "wave-like contour" (see Figure 3), usually to reinforce the expressive message in exclamative and imperative sentences.

4) **Variations in F0 range** of whole sentence, associated to the expression of emotions by speakers.

TABLE A: Mean values of Slope (in Hz/ms.) and F0 Height (in Hz relative to the initial F0 value) in main trace segments, calculated for each sentence type

Sencence type	Final Slope	FIRST RISING SEGMT.		FINAL SEGMENT	
		Slope	F0 Height	Slope	F0 Height
Enunciative	-	0,38	59	-0'27	-29
Int. Type I	+	0,65	96	0'87	171
Int. Type II	+	0,57	123	1'06	146
	-	0,75	80	-0'52	-42
Int. Type III	+	0,70	110	0'79	157
Exclamative	+	0,72	119	0'48	13
	-	0,51	93	-0'49	-35
Imperative	+	0,36	43	0'40	15
	-	0,61	59	-0'42	-38

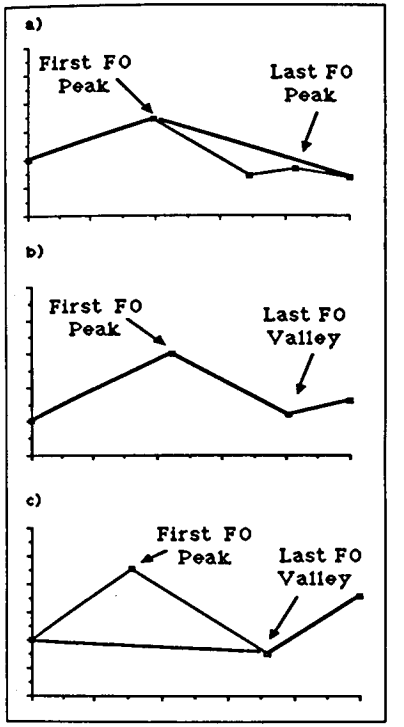

FIGURE 2: Typical F0 patterns in Spanish: a) Falling; b) Slow Rising; c) Fast Rising.

4. APPLICATION OF RESULTS TO AUTOMATIC RECOGNITION

This stylization method could be adapted to automatic implementation. A **smoothing procedure** would be introduced, to eliminate variations due to F0 detection errors and to interpolate segments were F0 is not present, and a different time normalization procedure would be used, since the one applied at this study is not useful with varying length sentences.

The obtained models are being applied to automatic labelling of sencence types. Slope and relative height of initial and final segments of stylized contours seem sufficient cues for the identification of different types.

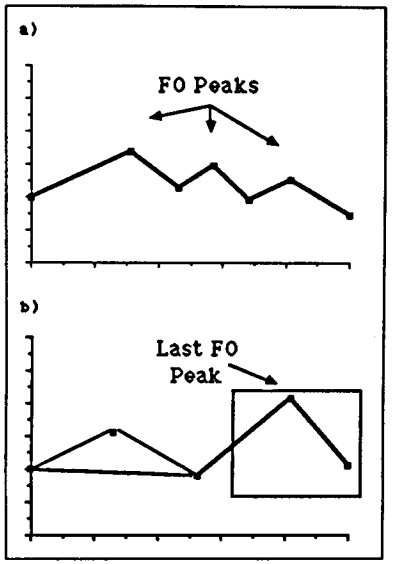

FIGURE 3: Special F0 patterns: a) "Wave-like" Contour; b) Circumflex Final.

Further research to verify the perceptual salience of this schematic models and its validity in continuous speech sentences has still to be carried out .

This work has been supported by a grant of the CIRIT from the Catalan Government.

REFERENCES

[1] MERTENS, P. (1989), "Automatic Recognition of Intonation in French and Dutch", in TUBACH, J.P. - MARIANI, J.J. (Eds.), *Eurospeech 89*, Vol. I, Paris, 46-50.

[2] NAVARRO TOMAS, T. (1948), *Manual de entonación española*, Madrid: Guadarrama.

[3] PIERREHUMBERT, J. (1987), *The Phonology and Phonetics of English Intonation*, Bloomington: Indiana University Linguistics Club.

[4] VAISSIÈRE, J. (1989), "On Automatic Extraction of Prosodic Information for Automatic Speech Recognition Systems", in TUBACH, J.P. - MARIANI, J.J. (Eds.), *Eurospeech 89*, Vol. I, Paris, 202-205.

[5] TAKEFUTA, Y. (1975), "Method of Acoustic Analysis of Intonation", in SINGH, S. (Ed.), *Measurement Procedures in Speech, Hearing and Language*, Baltimore: University Park Press, 173-186.