

# UNA APROXIMACIÓN FONÉTICA AL ESTUDIO DE LA ENTONACIÓN

*Mònica Estruch, Juan M. Garrido, Joaquim Llisterri, Montse Riera*

This article presents a methodology for the phonetic analysis of intonation contours, and the results of its application to the study of Catalan and Spanish intonation by the *Grup de Fonètica* of the *Universitat Autònoma de Barcelona* (UAB). The methodology presented here is based on a similar approach developed by the IPO (Eindhoven, the Netherlands), and includes four different tasks. The first task involves the extraction of F0 contours by means of a pitch tracker. However, the output of these systems contains estimation errors and micromelodic variations. A stylization procedure is applied to the F0 contours to eliminate these variations of the contour representation. Several stylization procedures are presented, and the main advantages and disadvantages of each one are outlined. The next task involves the definition of intonation patterns from the obtained stylized contours. The results of several studies carried out by the *Grup de Fonètica*, concerning the definition of local (at the stress group level) and global patterns (at the intonation group, sentence or paragraph level) for Spanish and Catalan, are summarized. Once the intonation patterns have been defined, the next task analyzes the use of the defined patterns in the transmission of linguistic (stress, syntactic structure, etc.) and non-linguistic information (dialect, emotions, or social origin of the speaker). Several works related to this topic carried out by the group are also outlined. The final goal of this methodology is the definition of intonation models for Spanish and Catalan, and their use in text-to-speech applications.

## 1. Introducción

El estudio de los elementos suprasegmentales o prosódicos constituye uno de los ámbitos de la fonética que ha recibido una mayor atención en los últimos años<sup>1</sup>, no únicamente por su interés teórico y descriptivo, sino también por sus implicaciones en lo que respecta a las mejoras en la naturalidad de los sistemas de conversión de texto a habla que pueden obtenerse mediante una modelización prosódica lo más cercana posible a la del habla natural. El presente trabajo constituye una contribución al desarrollo de una metodología para el estudio de la entonación<sup>2</sup> ejemplificada con algunos de los resultados obtenidos en el análisis del español y del catalán, en el contexto de un proyecto más amplio orientado hacia la formalización de conocimientos fonéticos que puedan ser aplicados tanto a la descripción lingüística como al desarrollo de tecnologías del habla.

---

<sup>1</sup> Véanse, por ejemplo, las actas del simposio organizado por ESCA (*European Speech Communication Association*) en Lund, Suecia, (House y Touati, Eds, 1993) y las sesiones dedicadas a la prosodia en el reciente congreso de la misma asociación celebrado en Madrid (Pardo *et al.* Eds., 1995).

<sup>2</sup> Utilizaremos el término 'entonación' para referirnos a la evolución de la curva melódica a lo largo de los enunciados.

Esta metodología, desarrollada a partir de 1990 y descrita principalmente en Garrido (1996), se enmarca en las aproximaciones fonéticas al estudio de la entonación en el sentido de que, partiendo de la representación física de las curvas melódicas, se pretende llegar a determinar su valor fonológico (o funcional). Está basada principalmente en el modelo formulado en el IPO (Eindhoven, Holanda), que se ha aplicado al estudio del holandés y de otras lenguas ('t Hart - Collier - Cohen, 1990).

La aproximación fonética propuesta aquí comprende cuatro etapas:

1. Extracción de las curvas melódicas.
2. Estilización.
3. Definición de patrones melódicos.
4. Estudio de la relación entre patrones e información lingüística, sociolingüística y estilística.

En los apartados siguientes se describen con más detalle las tres últimas etapas, y se presentan brevemente los resultados obtenidos hasta el momento en cada una de ellas.

## 2. Estilización de curvas melódicas

Las curvas melódicas<sup>3</sup>, tal como se obtienen a partir de los algoritmos de extracción de frecuencia fundamental (F0), plantean problemas para su utilización en el análisis de la entonación. En primer lugar, los detectores de F0 cometen a menudo errores, que dificultan la interpretación de las curvas, y que es necesario eliminar. En segundo lugar, las curvas melódicas contienen variaciones y rupturas, relacionadas con las características de los elementos segmentales que componen los enunciados (interrupciones en la curva debidas a la presencia de segmentos sordos, pequeñas variaciones en el curso de la curva melódica debidas a la naturaleza de los elementos segmentales). Estas variaciones, llamadas micromelódicas, no aportan ningún tipo de información lingüística (aunque pueden ser un indicio adicional para el reconocimiento del segmento en cuestión), e incluso puede que no sean percibidas por el oyente. Por tanto, parece conveniente no tenerlas en cuenta a la hora de interpretar la configuración de las curvas melódicas en términos funcionales.

La estilización y la codificación son los procedimientos que se usan en general para solventar estos problemas. La codificación implica normalmente una aproximación más fonológica al estudio de la entonación<sup>4</sup>, mientras que la estilización es un procedimiento más ligado a aproximaciones fonéticas. En la estilización, las curvas melódicas quedan reducidas a una serie discreta de puntos que, unidos mediante líneas, proporcionan una representación de sus movimientos más relevantes. La estilización puede llevarse a cabo de diferentes maneras. A continuación se describen algunas de las que se han venido utilizando en el *Grup de Fonètica* del *Departament de Filologia Espanyola* de la UAB para el análisis de la entonación.

<sup>3</sup> Por curva melódica o contorno melódico se entiende aquí la evolución temporal de la frecuencia fundamental.

<sup>4</sup> Para una revisión de los principales esquemas de codificación prosódica véase Llisterri (1994).

### 2.1. Métodos manuales de estilización

El proceso de estilización puede consistir en la medida del valor de la F0 en unos puntos determinados de la curva, con objeto de obtener una versión estilizada de la misma, de la que se han eliminado (al menos parcialmente) las variaciones micromelódicas. En Garrido (1991) se definió un primer sistema de estilización manual, basado en las diferencias frecuenciales entre los diferentes puntos de la curva melódica. Dicho sistema fue empleado para el análisis de un corpus de frases aisladas del español que contenían únicamente segmentos sonoros.

Sin embargo, cuando lo que se pretende es el análisis de materiales menos controlados, dicho sistema de análisis resulta poco adecuado, debido principalmente a la presencia de las variaciones micromelódicas. Dado que los segmentos vocálicos son las partes que se ven menos afectadas por las variaciones de este tipo, los valores se pueden tomar en algún (o algunos) punto(s) de la curva melódica coincidentes con las vocales del enunciado analizado. Los valores pueden obtenerse, por ejemplo, determinando la F0 en el centro de las vocales que constituyen el núcleo silábico, como en Garrido *et al.* (1993, 1995). Sin embargo, hay variaciones de F0 que se producen dentro de una misma vocal que con este método pueden no quedar reflejadas en la estilización resultante. Por ello, se han realizado estudios que consideran más de un valor de F0 para cada vocal, como en Estruch - Garrido (1995), en el que la representación de las curvas se obtuvo partiendo del valor de la F0 en el inicio, centro y final de la vocal, tal como ilustra la figura 1.

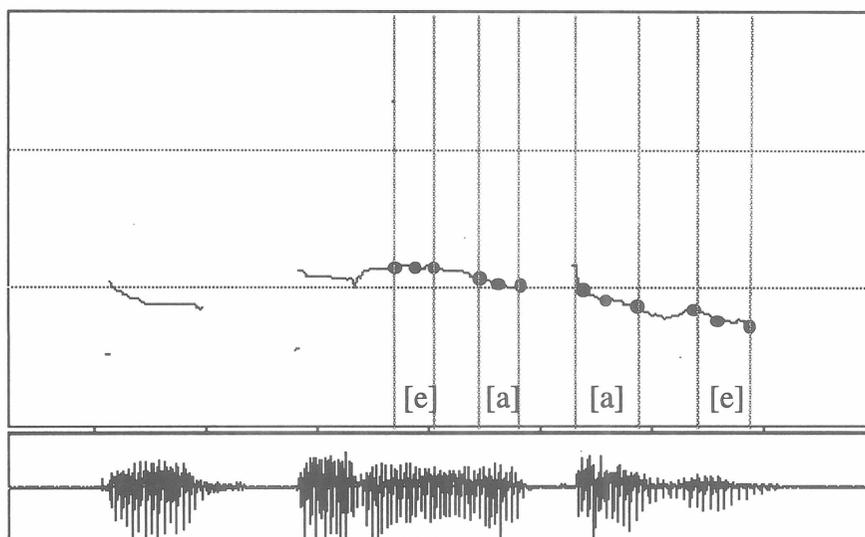


Figura 1. Curva melódica correspondiente al enunciado 'Te espero en la calle'. Las líneas verticales indican los límites de las vocales del grupo acentual 'en la calle'. Los puntos de cada vocal en que se tomaron los valores de F0 están marcados con círculos negros.

Estos métodos resultan adecuados para el análisis de corpus no demasiado extensos, como en el caso de los estudios mencionados. Para la estilización de curvas en un corpus más amplio, es preferible, sin embargo, el uso de un sistema automático.

## 2.2. Métodos automáticos de estilización

Los métodos automáticos, al igual que los manuales, convierten la curva melódica en una serie de puntos de inflexión<sup>5</sup> unidos mediante líneas. Las diferencias entre los diferentes métodos radican en el procedimiento de obtención de los puntos de inflexión, por un lado, y en el sistema de interpolación de los puntos por otro.

En el marco del proyecto LRE MULTEXT (*Multilingual Text Tools and Corpora*) (Hirst *et al.*, 1994) se ha utilizado el sistema MOMEL, un procedimiento de estilización automática desarrollado en el *Laboratoire Parole et Langage* (CNRS, *Université de Provence*) en Aix-en-Provence (Francia). El sistema determina los puntos de inflexión (*target points*) por medio de un cálculo de regresión que aproxima la forma de la curva de F0 y elimina las variaciones irrelevantes (Hirst - Espesser, 1993). Este sistema ha sido empleado en la UAB para anotar prosódicamente una parte del corpus EUROM1 (Chan *et al.*, 1995) del español y del francés.

Otro sistema de estilización empleado ha sido desarrollado por miembros del CITVAM de *Enginyeria La Salle* de la *Universitat Ramon Llull* de Barcelona, y está integrado en el programa 'Pitch', una aplicación para Windows que permite la obtención, estilización y modificación de la curva melódica de los enunciados, así como su posterior síntesis para la evaluación perceptiva de las estilizaciones obtenidas (Martínez, 1995). Este sistema de estilización determina automáticamente los puntos de inflexión en función de un 'umbral de semejanza' establecido previamente por el usuario, de manera que éste puede definir hasta qué punto la curva estilizada obtenida se asemejará a la curva original. Este sistema ha sido utilizado, entre otros trabajos, en Estruch *et al.* (1995) y Garrido (1996). La figura 2 presenta un ejemplo de curva estilizada obtenida mediante el programa 'Pitch'.

---

<sup>5</sup> Los puntos de inflexión son aquellos puntos de la curva melódica en los que se produce un cambio en el valor de la F0 que afecta de manera significativa a la evolución de la curva.

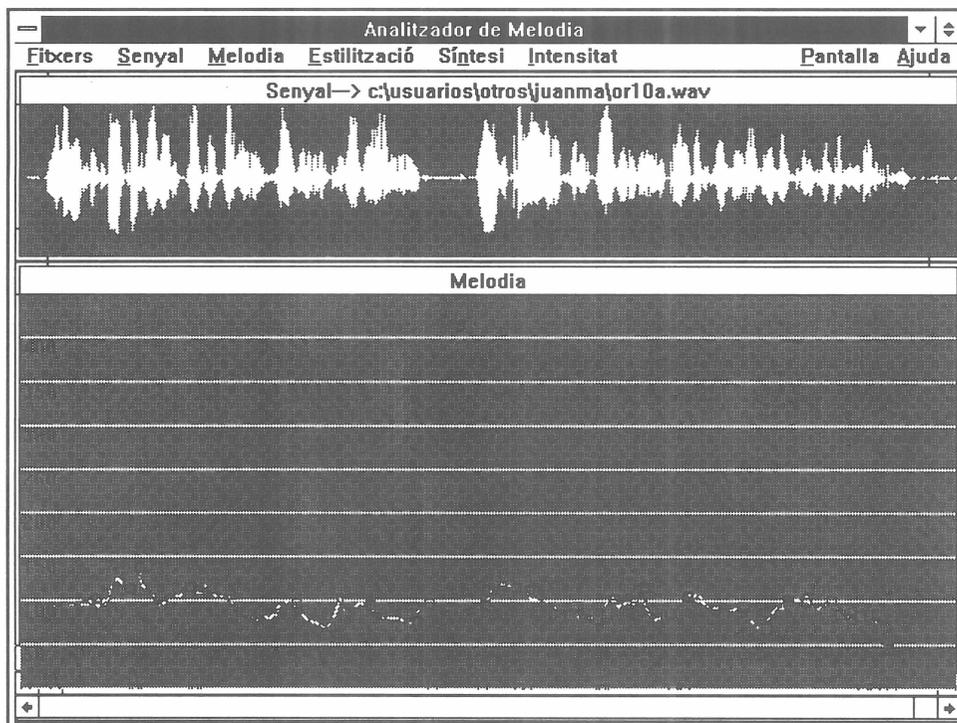


Figura 2. Curva melódica original y contorno estilizado del enunciado *'La organización terrorista ha protagonizado en lo que va de año una escalada de atentados en los que han perdido la vida quince personas'*, obtenidos mediante el programa 'Pitch'.

Una de las tareas previas a la utilización de este sistema ha sido la definición del 'umbral de semejanza' que debía emplearse para obtener curvas estilizadas idénticas perceptivamente a las correspondientes curvas originales. Este umbral de semejanza se definió experimentalmente por medio de pruebas de percepción, descritas en Jiménez (1994) y Estruch *et al.* (1995). El análisis de los resultados llevó a definir el valor del 8% como el más adecuado para obtener curvas melódicas equivalentes perceptivamente a la original, y con el menor número posible de puntos de inflexión.

### 3. Análisis de los patrones melódicos

Una vez obtenida una representación estilizada de las curvas melódicas, el siguiente paso es definir patrones melódicos<sup>6</sup> a partir de dichas representaciones.

<sup>6</sup> Por patrón melódico se entiende una representación estilizada de la curva melódica que recoge las características comunes a un conjunto de estilizaciones que comparten determinadas propiedades fonéticas y/o lingüísticas.

Los patrones melódicos pueden dividirse, siguiendo la clasificación propuesta en Garrido (1996), en dos grandes tipos: los patrones locales, que serían patrones de ámbito reducido (uno o dos grupos acentuales como máximo), y los patrones globales, de ámbito más extenso (grupos melódicos, frases o incluso párrafos), a los cuales se superpondrían los patrones locales para componer la curva melódica final.

### 3.1. Patrones locales

Una manera de representar los patrones locales a partir de contornos estilizados es en forma de series de puntos de inflexión unidos por movimientos descendentes, ascendentes y planos. Este tipo de representación es válida tanto para los movimientos en posición final de grupo fónico (patrones finales) como para los movimientos en posición inicial o interior de grupo (patrones no finales).

#### 3.1.1. Movimientos no finales

La forma típica de los patrones no finales, de acuerdo con la descripción presentada en Garrido (1996), consiste en una primera fase de descenso de la F0, seguida de una fase de ascenso. El punto de inflexión que marca el fin de la fase de descenso y el inicio de la fase de ascenso suele situarse al principio de la sílaba tónica del grupo acentual. Sin embargo, la posición del punto de inflexión que marca el final del movimiento ascendente (el 'pico') parece mucho menos estable: los resultados presentados en Garrido *et al.* (1993, 1995) y Llisterri *et al.* (1995) muestran un fenómeno de desplazamiento del pico hacia la sílaba posterior a la acentuada —ampliamente documentado también en otras lenguas—, que parece depender de factores tales como el tipo de límite sintáctico que se encuentra a continuación.

#### 3.1.2. Movimientos finales

Las descripciones tradicionales de la entonación del español han centrado su atención en la forma de los movimientos finales. La presentación clásica de Navarro Tomás (1944), por ejemplo, propone una clasificación de los contornos finales en cinco tonemas con valor distintivo: anticadencia, semianticadencia, suspensión, semicadencia y cadencia. Sin embargo, algunos trabajos llevados a cabo tanto desde el punto de vista acústico (Estruch - Garrido, 1995) como perceptivo (Estruch - Riera, 1996a) parecen indicar que esta clasificación podría ser revisada.

A la vista de los resultados experimentales, en Estruch - Garrido (1995) se ha propuesto una clasificación alternativa de los movimientos finales en español, en la que los 17 patrones melódicos finales definidos están agrupados en cuatro categorías: descendente, ascendente-descendente, ascendente y descendente-ascendente. La figura 3 proporciona un ejemplo de cada una. Una clasificación semejante se ha propuesto para los patrones finales del catalán en Estruch (1996).

Una segunda clasificación de los patrones finales del español, incluida en Garrido (1996), propone sólo tres categorías: descendentes, ascendentes y circunflejos.

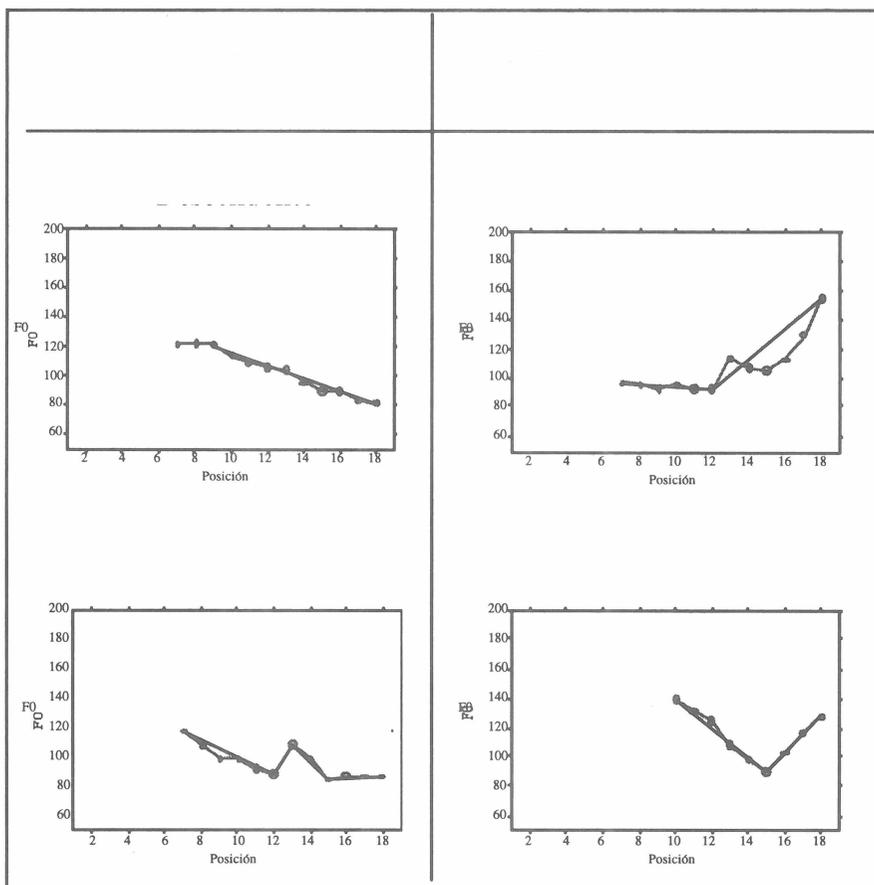


Figura 3: Clasificación de los patrones finales del español definidos en Estruch-Garrido (1995)

### 3.2. Patrones globales

La definición clásica de la entonación relaciona ésta con la forma global de la curva melódica a lo largo de un grupo fónico, es decir, con su patrón global. Los patrones definidos por Navarro Tomás (1944) para la frase simple en español se basan en esta idea. Siguiendo esta línea, se han propuesto inventarios de patrones melódicos de la frase simple para el español (Garrido, 1991) y el catalán (Estruch *et al.*, 1994).

Sin embargo, los resultados de una prueba de percepción presentados en Estruch *et al.* (1995) parecen indicar que este tipo de patrones son válidos para frases relativamente cortas pero no para modelar la entonación de frases más largas. Otros métodos de representación,

como el basado en el concepto de las líneas de referencia, o líneas de declinación<sup>7</sup>, empleado para la representación de los patrones globales en otras lenguas, aparecen como una alternativa interesante al planteamiento tradicional. Este sistema ha sido aplicado al español en Garrido *et al.* (1993, 1995) y Garrido (1996).

### 3.2.1. Aplicación del sistema de Navarro Tomás a la definición de los patrones melódicos de la frase simple en español y catalán

El procedimiento utilizado en Garrido (1991) y Estruch *et al.* (1994) se basa en el principio de que los patrones finales son el resultado de agrupar las estilizaciones cuya forma se asemeja y definir un patrón tipo con los puntos de inflexión que están presentes en todos (o en la mayoría) de los contornos estilizados del grupo. Los patrones obtenidos mediante este procedimiento tienen la apariencia de los presentados en la figura 4, definidos para el catalán en Estruch *et al.* (1994).

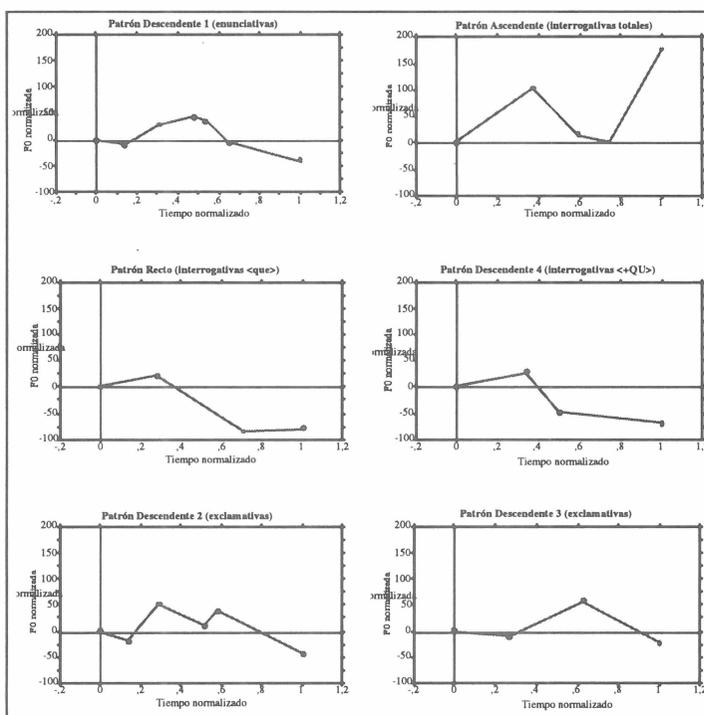


Figura 4: Patrones melódicos de la frase simple del catalán definidos en Estruch *et al.* (1994)

<sup>7</sup> La declinación es un fenómeno consistente en el descenso progresivo de la frecuencia fundamental a lo largo de un enunciado. Este fenómeno puede representarse mediante las llamadas líneas de referencia que establecen los márgenes frecuenciales dentro de los cuales pueden oscilar los valores de F0 a lo largo de un enunciado.

## 3.2.2. Aplicación del concepto de líneas de referencia a la entonación del español

En Garrido (1996) se presenta una representación de los patrones del grupo fónico basada en el uso de líneas de declinación. Se proponen tres líneas de referencia que definen tres niveles tonales diferentes (superior, medio e inferior) sobre los que se situarían los diferentes puntos de inflexión etiquetados como P (pico), M (medio) y V (valle) respectivamente. La figura 5 ilustra este procedimiento.

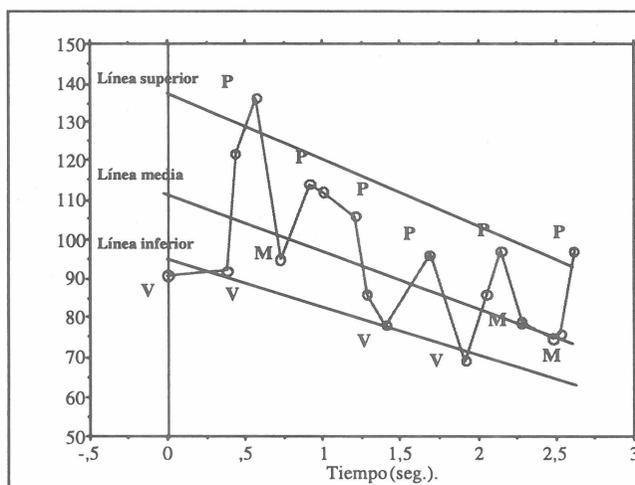


Figura 5. Líneas superior, media e inferior correspondientes al enunciado 'La organización terrorista ha protagonizado en lo que va de año', pronunciado por un locutor masculino.

Una ventaja del sistema de representación de los patrones globales mediante líneas de referencia es que puede aplicarse también a la representación de fenómenos de declinación que se dan en unidades superiores como el párrafo (el llamado fenómeno de la supradclinación). En Garrido (1993), Garrido *et al.* (1993, 1995) y Garrido (1996) se analiza la existencia de este tipo de supradclinación en español, y se avanza la posibilidad de que pueda representarse mediante 'supra-líneas' de declinación como las que aparecen en la figura 6.

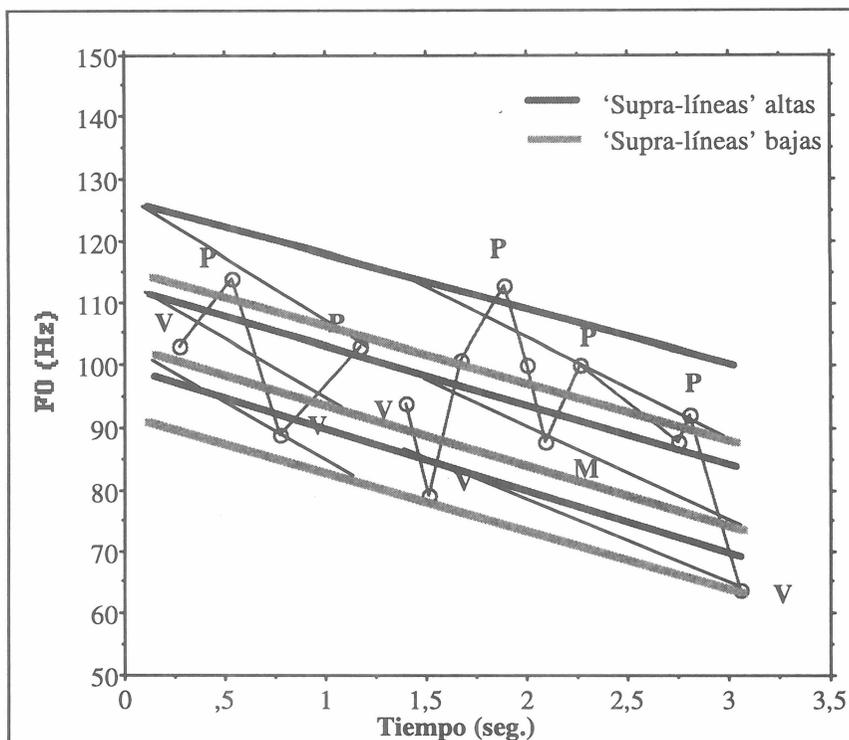


Figura 6. 'Supra-líneas' superior e inferior calculadas para los grupos fónicos del enunciado 'La organización terrorista ha protagonizado en lo que va de año una escalada de atentados en los que han perdido la vida quince personas', pronunciado por un locutor masculino.

#### 4. Usos de los patrones melódicos

Tras determinar los patrones melódicos, el siguiente paso en el estudio de la entonación es determinar cuál es la información lingüística (acento, estructura sintáctica y pragmática), paralingüística (procedencia geográfica y social, estados emocionales, etc.) que el hablante transmite al usar dichos patrones, o, expresado de otra manera, cuál es el efecto perceptivo que éstos provocan en el oyente.

Por lo que respecta a la relación entre entonación e información lingüística, la modalidad oracional es el aspecto más estudiado en español. Los trabajos clásicos como el de Navarro Tomás (1944) han descrito en detalle la expresión de la modalidad oracional por medio de patrones entonativos. Estos resultados han sido confirmados posteriormente por los obtenidos en estudios como Garrido (1991) sobre los patrones globales o Estruch - Garrido (1995) sobre los patrones locales finales. Por lo que respecta al catalán, los trabajos de Estruch *et al.* (1994) y Estruch *et al.* (1995) ofrecen una descripción de la relación entre los patrones globales y la modalidad oracional.

En cambio, la relación entre entonación y estructura sintáctica ha sido menos analizada en español. Trabajos como los de Garrido *et al.* (1995) y Llisterri *et al.* (1995) han explorado las relaciones entre los patrones locales, finales y no finales, y los límites sintácticos.

Finalmente, la relación entre entonación e información estilística ha comenzado a estudiarse en Estruch - Riera (1996b), aunque los resultados pueden considerarse todavía preliminares.

### **5. Objetivo final: un modelo jerárquico para la entonación del español y el catalán**

El objetivo último de esta metodología es definir un modelo de uso de los patrones melódicos del español y del catalán, que permita predecir la combinación de los diferentes patrones en la curva melódica final en función de una serie de factores lingüísticos, sociolingüísticos o estilísticos. En Garrido (1996) se hace una primera propuesta de modelo melódico para el español en lectura, que comprende dos niveles entonativos: un nivel local y un nivel global. En el nivel global se definirían los patrones globales más adecuados para el enunciado, representados en forma de líneas de declinación. A estos patrones globales se superpodrían en un segundo paso los patrones locales, que definirían la forma final de la curva melódica.

Este modelo tiene como aplicación más inmediata la descripción fonética de la entonación en un marco teórico comparable al utilizado para otras lenguas, a la vez que permite elaborar un modelo entonativo de utilidad en el desarrollo de sistemas de conversión de texto a habla.

**BIBLIOGRAFÍA**

- D. Chan - A. Fourcin - D. Gibbon - B. Granström - M. Hucvale - G. Kokkinaki - K. Kvale - L. Lamel - B. Lindberg - A. Moreno - J. Mouropoulos - F. Senia - I. Trancoso - C. Veld - J. Zeiliger, «EUROM- A Spoken Language Resource for the EU», in *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology* (Madrid 18-21 September, 1995) Vol 1, 867-870.
- M. Estruch, *Anàlisi i classificació dels contorns melòdics finals en un corpus de frases aïllades del català* (Bellaterra 1996).
- M. Estruch - S. López - M. Mercadal - M. Riera, *Comparació dels patrons melòdics del català i del castellà* (Bellaterra 1994).
- M. Estruch - M. Riera, *Validació perceptiva del model entonatiu de Navarro Tomás* (Bellaterra 1996a).
- M. Estruch - M. Riera, *Validació perceptiva de quatre emocions del català* (Bellaterra 1996b).
- M. Estruch - J. M. Garrido, «Análisis y clasificación de los contornos melódicos finales en un corpus de frases aisladas del español», *Revista Española de Lingüística*, 26,1 (1995) 138-139.
- M. Estruch, M. Mimó - J. Renom - M. Riera, *Validació perceptiva dels patrons melòdics del català* (Bellaterra 1995).
- J.M. Garrido, *Modelización de patrones entonativos para la síntesis y el reconocimiento de habla* (Bellaterra 1991).
- J.M. Garrido, «Analysis of Global Pitch Contour Domains at Paragraph Level in Spanish Reading Text» in *Proceedings of an ESCA Workshop on Prosody* (Lund 1993) 104-107.
- J.M. Garrido, *Modelling Spanish Intonation for Text-to-Speech Applications* (Bellaterra 1996).
- J.M. Garrido - J. Llisterra - R. Marín - C. De La Mota - A. Ríos, «Prosodic markers at syntactic boundaries in Spanish», in *Proceedings of the XIIIth International Congress of Phonetic Sciences*. (Stockholm 13-19 august 1994) Vol 2, 370-373.
- J.M. Garrido - J. Llisterra - C. De La Mota - A. Ríos, «Prosodic differences in reading style: Isolated vs. Contextualized Sentences», in *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. (Berlin 21-23 September 1993) Vol 1, 573-576.
- J. M. Garrido - J. Llisterra - C. De La Mota - A. Ríos, «Estudio comparado de las características prosódicas de la oración simple en español en dos modalidades de lectura», en A. Elejabeitia - A. Iribar (Eds.) *Phonetica. Trabajos de fonética experimental*. (Bilbao 1995) 173-194.
- D. J. Hirst - R. Espesser, «Automatic modelling of fundamental frequency using a quadratic spline function», *Travaux de l'Institut de Phonétique d'Aix*, 15 (1993) 71-85.

- D. J. Hirst - N. Ide - J. Véronis, «Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTTEXT project», in *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*. (New York September 12-15, 1994) 77-80.
- D. House - P. Touati (Eds), *Proceedings of an ESCA Workshop on Prosody*. (Lund, September 27-29, 1993).
- J. M. Jiménez, *Implementació d'un mètode d'estilitzat de corbes melòdiques* (Barcelona 1994).
- D. R. Ladd, «Declination 'reset' and the hierarchical organization of utterances», *Journal of the Acoustical Society of America* 84,2: (1988) 530-544.
- J. Llisterri, *Prosody Encoding Survey*. WP 1 Specifications and Standards. T1.5. Markup Specifications. Deliverable 1.5.3. Final version, (15 September 1994). LRE project 62-050 MULTTEXT.
- <http://www.lpl.univ-aix.fr/projects/multext/CES/CES2.html>
- J. Llisterri - R. Marín - C. De La Mota - A. Ríos, «Factors affecting F0 peak displacement in Spanish», in J. M. Pardo *et al.* (Eds) *Eurospeech'95 Proceedings. ESCA 4th European Conference on Speech Communication and Technology*. (Madrid, 18-21 September, 1995) Vol. 3, 2061-2064.
- D. Martínez, *Sistema d'anàlisi, tractament i síntesi de la melodia en entorn Windows* (Barcelona 1995).
- T. Navarro Tomás, *Manual de entonación española*, (Madrid 1944).
- J. M. Pardo - E. Enríquez - J. Ortega - J. Ferreiros - J. Macías - F. J. Valverde (Eds), *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology* (Madrid 18-21 September, 1995).
- J. 'T Hart - R. Collier - A. Cohen, *A perceptual study of intonation. An experimental-phonetic approach to speech melody* (Cambridge 1990).

