

5. MODELO DE AIX-EN-PROVENCE

Lorraine BAQUÉ y Mònica ESTRUCH
(Universitat Autònoma de Barcelona)

5.1. Introducción

El modelo de Aix-en-Provence se ha venido desarrollando en el Laboratorio Parole et Langage de la Universidad de Aix-en-Provence (véase www.lpl.univ-aix.fr) por Daniel J. Hirst, Albert Di Cristo y Robert Espesser durante estos últimos años. Pretende obtener una modelización fonológica del sistema prosódico de distintas lenguas con el fin de dar cuenta de las características prosódicas de cada una de esas lenguas, y además llevar a cabo estudios de comparativismo prosódico para extraer aquellos aspectos que son invariantes o universales y aquellos que son específicos de un sistema lingüístico determinado (véase Hirst y Di Cristo, 1998).

Se enraíza en la línea del generativismo y de las teorías fonológicas¹ de la entonación. En efecto, por una parte, los autores, como suele ser usual desde los inicios del generativismo, distinguen en todos los ámbitos de la lingüística —y en particular en la prosodia— dos niveles básicos:

- un nivel de representación lingüística abstracto (nivel cognitivo) que da cuenta de la competencia lingüística del locutor nativo, esto es, de los conocimientos que debe adquirir un locutor que aprende la lengua, y
- un nivel de representación analítico que permite dar cuenta de la realización concreta de los conocimientos lingüísticos.

Asimismo, Hirst, Di Cristo y Espesser (2000) consideran que la prosodia es parte integrante del componente fonológico de la lengua, es decir, que, por el “principio de autonomía”, constituye un módulo independiente del sentido que transmite, ya que se modeliza el lenguaje de manera secuencial de tal forma que el módulo fonológico constituye la última etapa del nivel cognitivo después de los módulos semántico y sintáctico.

Por otra parte, si asumimos que en todo modelo secuencial del lenguaje es preciso que se cumpla la “**condición de interpretabilidad**”, esto es, que cada nivel intermedio se pueda interpretar en los dos niveles adyacentes —el superior y el inferior—, la representación fonológica de la prosodia tiene un doble cometido:

- proporcionar la información necesaria para la interpretación sintáctica y semántica del enunciado, y
- proporcionar la información necesaria para pronunciarlo correctamente.

Con todo, si bien se postula que el componente fonológico interviene cuando la estructura semántica y sintáctica ya está programada y que cabe estudiarlo sin referirse

¹ Para una discusión de los enfoques fonológicos y no fonológicos de la entonación, véase [Rossi OSSI, 1999](#).

VERSIÓN BORRADOR PROVISIONAL

al sentido del enunciado, los autores prevén la posibilidad² de analizar posteriormente los lazos que existen entre la fonología y los componentes superiores.

Estos postulados básicos constituyen el concepto de lenguaje en que se apoya el modelo de Aix-en-Provence, pero no son, de hecho, ni exclusivos de dicho modelo, ni tampoco del componente prosódico.

Antes de entrar en una descripción más pormenorizada del modelo de Aix-en-Provence, es necesario comentar qué entienden sus autores por prosodia y cuál es el diagrama que mejor puede representar su estructura en un modelo de lenguaje.

En primer lugar, cabe señalar que en este enfoque la prosodia no se limita a la evolución de la curva melódica, sino que se trata de una estructura pluriparamétrica que se manifiesta mediante variaciones de frecuencia fundamental, intensidad y duración.

Por otro lado, se parte del supuesto de que las lenguas usan distintas formas prosódicas para codificar las mismas funciones prosódicas. Éstas competen a la teoría lingüística universal así como el inventario de las formas prosódicas, mientras que el modo según el cual se estructuran las funciones por medio de las distintas formas prosódicas varía de una lengua a otra y, por lo tanto, debe definirse mediante parámetros propios a cada sistema lingüístico. Se considera pues necesario no limitarse a la separación básica entre un nivel abstracto (fonológico) y un nivel concreto (fonético o físico), sino que cabe establecer otras distinciones en el seno de éstos.

Así, se establece una primera distinción entre el nivel físico (acústico o fisiológico) en el que se materializa cada enunciado concreto y el nivel fonético, basándose en la hipótesis de que todas las lenguas están sometidas a los mismos condicionantes de producción y percepción de las variaciones de frecuencia fundamental, intensidad y duración, y que, por consiguiente, una representación fonética es una estructuración particular de condicionantes universales.

Se establece una segunda distinción entre un nivel fonológico "superficial" que permite describir las diferencias interlingüísticas que se observan en las representaciones superficiales y el nivel fonológico "profundo", en el cual se distinguen asimismo las funciones y las formas prosódicas, a menudo confundidas en muchos de los sistemas de transcripción al uso.

En consecuencia, el modelo de Aix-en-Provence postula la existencia de cuatro niveles (véase Figura 1):

- un **nivel fonológico "profundo"** o "subyacente" que incluye las representaciones funcionales que codifican la información necesaria para la interpretación semántica y sintáctica de la prosodia de un enunciado,
- un **nivel fonológico "superficial"** en el cual se hallan las categorías discretas que permiten describir los fenómenos interlingüísticos de superficie,
- un **nivel fonético** que atañe a los fenómenos variables a partir de los cuales se han puesto de manifiesto los condicionantes universales de producción y percepción prosódicas³, y

² [Todavía no sistematizada en la versión actual del modelo. véase apartado 5.4.](#)

VERSIÓN BORRADOR PROVISIONAL

- un **nivel físico** que da cuenta de las características acústicas o fisiológicas de un enunciado.

Según la condición de interpretabilidad antes mencionada, es necesario que cada uno de estos niveles sea interpretable en el (o los) nivel(es) adyacente(s).

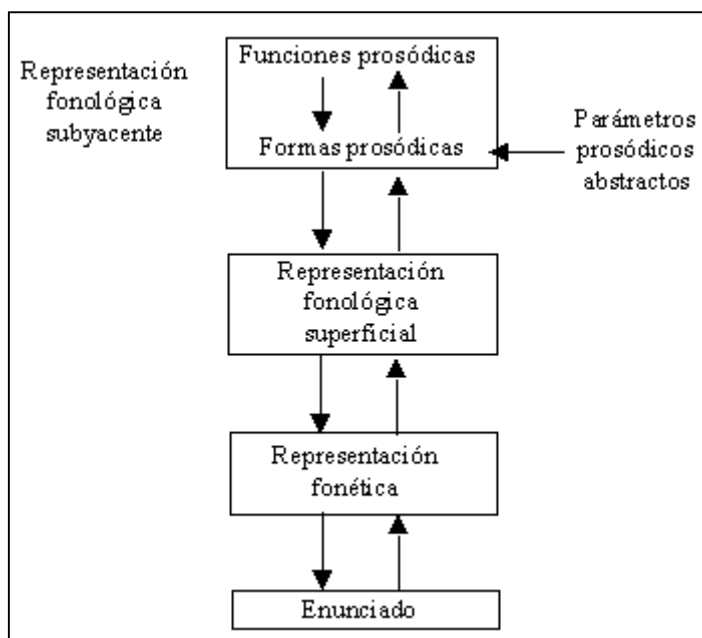


Figura 1. Niveles del módulo prosódico.

Así, las funciones prosódicas se estructuran en formas prosódicas subyacentes según parámetros prosódicos abstractos específicos de cada lengua. Estas formas se estructuran a su vez en las formas superficiales, y luego en las representaciones fonéticas, para llegar al final a los correlatos físicos de los enunciados.

Una vez definidos los niveles de representación en que se estructura el modelo de Aix-en-Provence, queda por definir cuál es el procedimiento que adoptan sus autores para acceder a una sistematización de los conocimientos prosódicos, y cuáles son los elementos que configuran las distintas representaciones.

Por una parte, este modelo comparte con las teorías de la fonología métrica y autosegmental⁴ el hecho de abogar, no por una configuración por **contornos entonativos** —que corresponde a la idea que la percepción de la entonación se basa fundamentalmente en los contrastes entre segmentos dentro del mismo grupo más que en la altura tonal que alcanza cada segmento—, sino por **segmentos tonales** —que implica que se concibe la entonación como una serie de elementos discretos que se yuxtaponen a lo largo del enunciado y que se atribuye una mayor importancia a la altura tonal de cada segmento que a las diferencias de tono entre uno y otro—, y postular dos niveles de representación básicos: los segmentos tonales discretos y una rejilla métrica con la correspondiente atribución de acentos.

³ Cf. Véase, a este respecto, Rossi OSSI et al. #, Di Cristo, Hirst, Martin y Nishinuma (1981).

⁴ Cf. Véase, a este respecto, el capítulo dedicado a ***6 de este libro.

VERSIÓN BORRADOR PROVISIONAL

En cuanto al método de investigación elegido, ya se ha mencionado que este modelo se inscribe en la línea de los herederos del generativismo, pero sus autores, contrariamente a otros con los que comparten algunos enfoques como Pierrehumbert⁵, están más influenciados por el estructuralismo europeo que por el americano, por lo que adoptan un razonamiento de índole deductiva, en el cual se atribuyen a priori unos tonos a las distintas unidades consideradas y es necesario, por consiguiente, definir **reglas de reajuste y linearización** (esto es, reglas que permitan transformar una cadena de tonos en una secuencia pronunciable) y criterios de correspondencia con los niveles inferiores para poder adaptar las unidades de la estructura subyacente a la realidad acústica observada.

Por otra parte, se trata de una **teoría linear jerárquica** en la medida en que:

- como en el modelo de Pierrehumbert, se concibe la entonación como una secuencia linear de tonos fonológicos que no interfieren con el contorno de grupo o de enunciado, y que
- contrariamente a lo que ocurre con la teoría linear de Pierrehumbert, por lo menos en un inicio⁶, la secuencia de tonos se estructura en tres unidades jerarquizadas de manera linear: la **unidad entonativa** (UE) entendida generalmente como la unidad comprendida entre dos pausas, la unidad rítmica (UR) y la **unidad tonal** (UT) definida como la agrupación de una sílaba tónica y las átonas de su alrededor. Cabe señalar que la **unidad rítmica** —intermedia entre la UE y la UT, caracterizada por contener un único acento léxico y una o más UT— no aparecía en las primeras versiones del modelo, sino que se añadió posteriormente para dar cuenta de la predominancia del acento léxico respecto del acento inicial, y así tomar en cuenta a la vez aspectos temporales (accento léxico) y melódicos (accento inicial)⁷.

Así, si recapitulamos lo anteriormente visto, un primer nivel de concreción del modelo de Aix-en-Provence queda definido de tal manera que el nivel fonológico profundo que, como hemos visto, pretende dar cuenta de la competencia del locutor, y que pertenece al nivel de representación cognitiva como la sintaxis y la semántica, tiene el cometido de generar patrones entonativos abstractos independientes del contenido lingüístico mediante una estructura abstracta organizada de manera jerárquica y asociada a plantillas tonales construidas a partir de dos tonos primitivos (H y L) sobre una rejilla métrica en la cual existen actualmente seis categorías no terminales⁸:

- la sílaba, asociada a la salida de la rejilla métrica,
- la unidad tonal, definida en función de una secuencia tonal atribuida a priori según la lengua considerada,
- la unidad acentuable,

⁵ Cf. Véase, a este respecto, el capítulo dedicado a ***6 de este libro.

⁶ En efecto, en las últimas versiones de la teoría de Pierrehumbert, desarrolladas conjuntamente con Beckman a partir de 1986, se introduce la noción de grupo intermedio que tiende a proporcionar a su teoría linear una dimensión jerárquica (véase Pierrehumbert y Beckman, 1988).

⁷ Otra diferencia con la teoría linear de Pierrehumbert es que ésta asimila la entonación al parámetro físico único de frecuencia fundamental, mientras que, como ya se ha señalado más arriba, el modelo de Aix-en-Provence considera que se trata de una estructura pluriparamétrica (frecuencia fundamental, intensidad y duración).

⁸ En el estado actual del modelo, algunas de estas categorías, postuladas a nivel teórico, precisan desarrollos ulteriores (cf. véase capítulo apartado 5.4).

VERSIÓN BORRADOR PROVISIONAL

- la unidad rítmica, definida, como la unidad tonal, en función de una secuencia tonal atribuida a priori según la lengua considerada,
- la unidad entonativa, definida en función de criterios sintácticos y/o acústicos, y
- la unidad de modulación, que representa el enunciado desde una perspectiva informativa.

Se atribuye a priori un tono H a todas las sílabas acentuadas, ya sea por la estructura del léxico o por razones de índole pragmática, y un tono L a todas las sílabas átonas; las unidades entonativas vienen limitadas por tonos de frontera (LH o LL) y a las unidades tonales corresponde una configuración tonal LH o HL en función de si la naturaleza de los constituyentes métricos es de tipo núcleo a la derecha o núcleo a la izquierda respectivamente.

El nivel fonético constituye una interfaz que permite establecer correspondencias entre el nivel acústico y los parámetros fonológicos, una vez resueltos los condicionantes de producción y percepción prosódica. Se parte de la hipótesis de que la entonación es la resultante de la superposición de dos componentes:

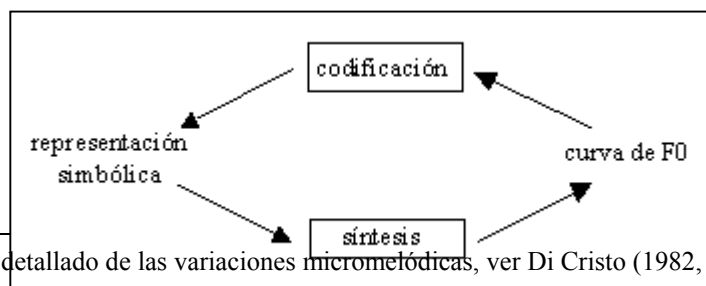
- un componente **microprosódico**⁹, que trata las variaciones prosódicas causadas por las características de los elementos segmentales que aparecen en el enunciado, y
- un componente **macroprosódico**, que refleja la selección que efectúa el hablante del patrón entonativo para la oración (lo que podríamos llamar la melodía de un enunciado),

por cuanto en este nivel se trata la curva melódica con el objetivo de separar aquellas características que pertenecen a uno u otro componente.

En cuanto al nivel fonológico superficial, constituye un módulo intermedio entre los dos anteriores. En él, las unidades de base no son movimientos de frecuencia fundamental, sino puntos estáticos interpretables como elementos discretos.

5.2. Componentes básicos del modelo

Existe un gran número de implementaciones de modelos fonéticos y fonológicos de entonación que han sido diseñados para derivar una salida acústica (la curva de F0) a partir de una entrada simbólica (serie de símbolos). Sin embargo, es mucho menos frecuente el planteamiento inverso: cómo recuperar una representación simbólica a partir de la curva de F0. El modelo de Aix-en-Provence plantea justamente esta interrelación entre la codificación y la síntesis, como se ejemplifica en la Figura 2.



⁹ Para un análisis detallado de las variaciones micromelódicas, ver Di Cristo (1982, 1985).

Figura 2. Interrelación entre la codificación y la síntesis en el modelo de Aix-en-Provence.

Estas dos grandes etapas (codificación y síntesis) se encuentran implementadas en el programa MES (*Motif Environnement for Speech*), desarrollado en el *Laboratoire Parole et Langage* de la Universidad de Aix-en-Provence. Este programa contiene un módulo de estilización automática (**MOMEL** – “MOdelling MELody”) y otro de anotación (**INTSINT** – “INternational Transcription System for INTonation”¹⁰); además, permite la alineación de los puntos con el texto transcrito ortográficamente y con los símbolos que se generan de manera automática con el sistema de codificación. La Figura 3 ejemplifica las diferentes fases del sistema.

¹⁰ Véanse, al respecto, [Hirst, Nicolas y Espesser \(1991\)](#); [Campione, Flachaire, Hirst y Véronis \(1997\)](#); [Campione, Hirst y Véronis \(2000\)](#); [Hirst y Di Cristo \(1998\)](#); [Hirst \(1999\)](#); [Hirst, Di Cristo y Espesser \(2000\)](#); así como el apartado que le está dedicado en el capítulo 8.

VERSIÓN BORRADOR PROVISIONAL

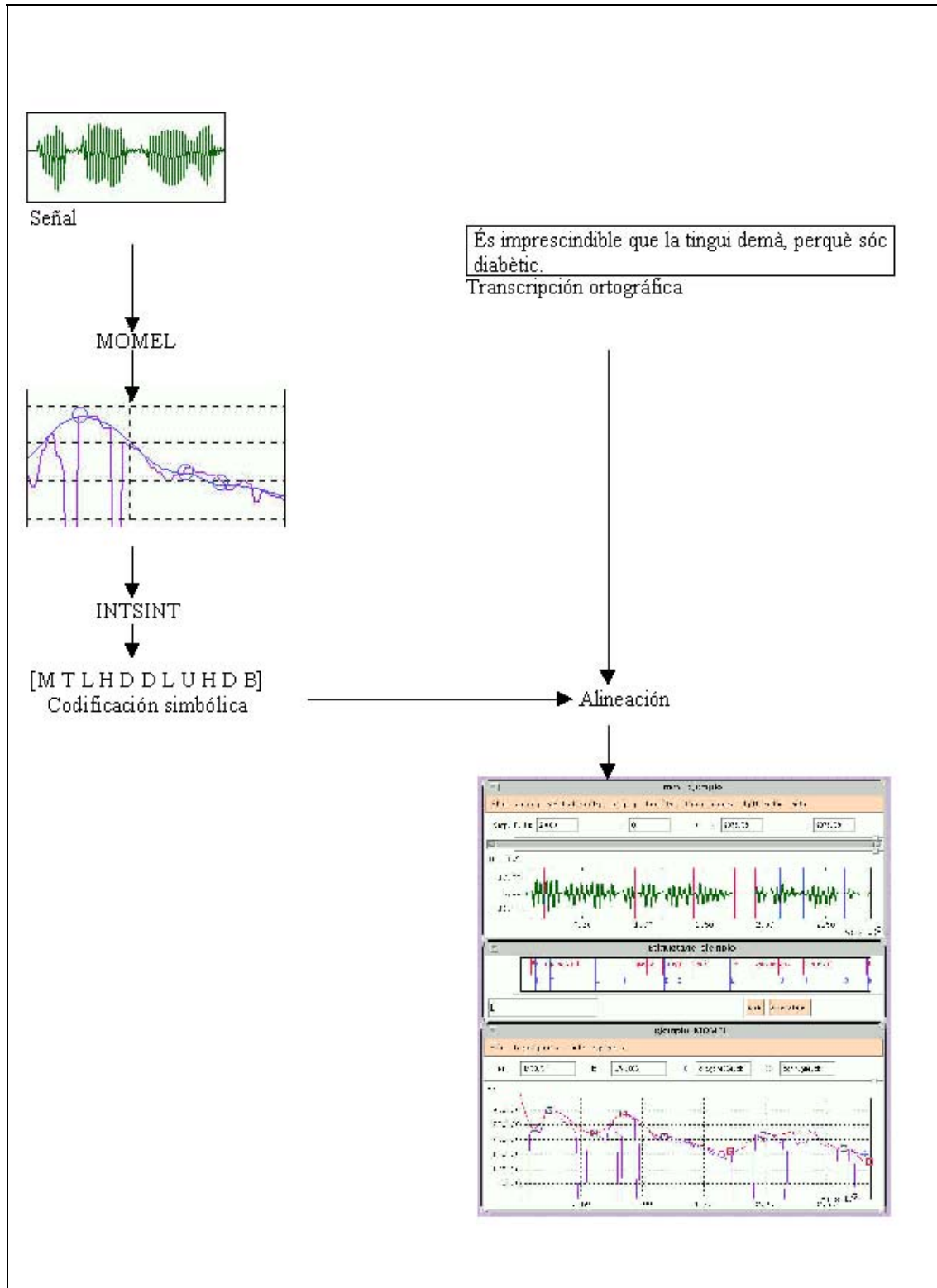


Figura 3. Esquema del proceso de tratamiento de la señal de voz con el programa MES.

El proceso ejemplificado en la Figura 3 es el siguiente: a partir de la señal de habla se deriva la curva de F0 y la representación parabólica a base de puntos (sistema MOMEL), que serán etiquetados en la fase siguiente (con el sistema INTSINT). Posteriormente, se alinea la señal con la curva de F0, los símbolos de la codificación, las etiquetas de límite de palabra, las etiquetas de inicio de sílaba tónica y la transcripción ortográfica.

Según los autores del modelo (Hirst, Di Cristo y Espesser, 2000), una teoría completa de la entonación necesita englobar cuatro niveles distintos de análisis, cada uno debiendo poder ser interpretado a partir del nivel precedente y/o del siguiente: un nivel físico (acústico y fisiológico), un nivel fonético, un nivel fonológico superficial y un nivel fonológico profundo. En consecuencia, el proceso ilustrado en la Figura 3 se estructura en estos cuatro niveles.

5.2.1. Nivel físico

La diferencia entre el nivel físico y el fonético, como se ha indicado ya en la introducción, se sustenta en la hipótesis de que todas las lenguas obedecen las mismas restricciones físicas en la producción y la percepción de las variaciones de la frecuencia del fundamental (F0), la intensidad y la duración. Por lo tanto, el nivel físico es aquel que podríamos etiquetar como universal, mientras que en el nivel fonético, estas restricciones universales ya no se tendrían en cuenta.

Así, por ejemplo, si partimos de la grabación de un enunciado catalán como el siguiente: *Es imprescindible que la tingui demà, perquè sóc diabètic.*¹¹, la primera fase consiste en extraer la curva de frecuencia fundamental de la señal acústica digitalizada a partir del método "vote"¹². Se obtiene así, además del oscilograma, una curva melódica previa a cualquier modelización.

5.2.2. Nivel fonético

Teniendo en cuenta que las curvas de F0 se modelan como la superposición de los componentes micro y macroprosódico, el objetivo de este nivel es tratar de extraer automáticamente la información macroprosódica relevante de la señal de habla. Esta extracción se concreta en dos fases (Campione, Hirst y Véronis, 2000):

- **estilización o modelización:** consiste en la sustitución de la curva de F0 por una función numérica sencilla que conserva la información macroprosódica original y deja de lado la información no relevante; se concreta en lo que se ha llamado la *close-copy stylisation* (de Pijper, 1979), esto es, una curva estilizada (simplificada) perceptivamente equivalente a la original.
- **codificación simbólica:** es la representación, mediante un alfabeto de símbolos, que convierte la curva estilizada en una secuencia de categorías discretas; estas categorías se pueden utilizar para regenerar la curva que, aunque perceptivamente no sea idéntica a la original, sí es percibida como equivalente lingüísticamente (de Pijper, 1979).

El algoritmo MOMEL permite estilizar automáticamente la curva de F0 como una secuencia de **puntos de inflexión** (*target points*) —esto es, aquellos puntos en que la curva modelizada sufre un cambio—, definiendo una función cuadrática (*quadratic spline function*) que une los puntos mediante parábolas (en lugar de las líneas rectas

¹¹ En [catalán castellano](#), significa: [:“Es imprescindible que la tenga mañana, porque soy diabético.”](#)

¹² [Para una descripción del método "vote", se consultará puede consultar el tutorial del programa en la página web \[http://www.lpl.univ-aix.fr/projects/mes_signaix.htm/mes_signaix.tutorial.html\]\(http://www.lpl.univ-aix.fr/projects/mes_signaix.htm/mes_signaix.tutorial.html\)](#).

VERSIÓN BORRADOR PROVISIONAL

usuales en otros modelos entonativos¹³). El resultado es una curva continua en la que incluso los segmentos sordos son interpolados para que no haya interrupciones en la curva, puesto que se trata de desligar la información segmental de la curva de F0. Este algoritmo consta de cuatro fases:

- Preprocesamiento de la frecuencia del fundamental: se eliminan algunos de los valores no relevantes de F0 de la curva.
- Estimación de los puntos de inflexión candidatos: se produce la neutralización de algunos valores que sobrepasan el límite máximo y mínimo de frecuencia (en función de los valores que conforman el rango del locutor) que podrían ser debidos a errores de detección de la frecuencia del fundamental; se aplica una regresión cuadrática a los otros valores; al final, a cada valor original de F0 se le asigna un valor de tiempo y de frecuencia o un valor nulo.
- Distribución de los puntos de inflexión candidatos: cada uno de los puntos no neutralizados en la fase anterior obtiene un valor de tiempo y de frecuencia.
- Reducción de los puntos de inflexión candidatos: se eliminan los puntos que presentan unos valores que sobrepasan la desviación estándar correspondiente.

Con este método se pretende obtener curvas estilizadas equivalentes perceptivamente a las originales, aunque puedan producirse algunos errores de detección de la frecuencia del fundamental que hay que corregir manualmente en una fase posterior. Este modelo de estilización ha sido utilizado y evaluado en diferentes lenguas, entre las cuales figuran el inglés, el francés, el español, el alemán, el sueco (Llisterri, 1996; Astesano, Espesser, Hirst y Llisterri, 1997) y el catalán (Estruch, 1999), con resultados globalmente positivos.

Retomemos el ejemplo de la frase catalana *És imprescindible que la tingui demà, perquè sóc diabètic*. A partir de los resultados obtenidos en la primera fase, con el algoritmo MOMEL se procede a la estilización automática que une mediante parábolas los puntos de inflexión obtenidos (marcados en la Figura 4 con el símbolo ○). De este modo, obtenemos una curva estilizada de la frecuencia fundamental que ha neutralizado los efectos del componente micromelódico. Posteriormente se evalúa auditivamente hasta qué punto la curva modelizada es perceptivamente equivalente a la curva original, y, eventualmente, se corrige manualmente la estilización automática añadiendo, suprimiendo o desplazando los puntos de inflexión que sea preciso. Se obtiene así, si es necesario, una curva melódica corregida (con puntos de inflexión marcados en la Figura 4 con el símbolo □). Al terminar dicho proceso, el resultado es el que hallamos en la figura siguiente:

¹³ Un estudio sobre las diferencias perceptivas entre puntos de inflexión interpolados por líneas rectas o parábolas se encuentra en t'Hart (1991). Campione, [Hirst y Véronis-et-al \(en prensa 2000\)](#) presentan argumentos a favor de la interpolación de los puntos mediante parábolas.

VERSIÓN BORRADOR PROVISIONAL

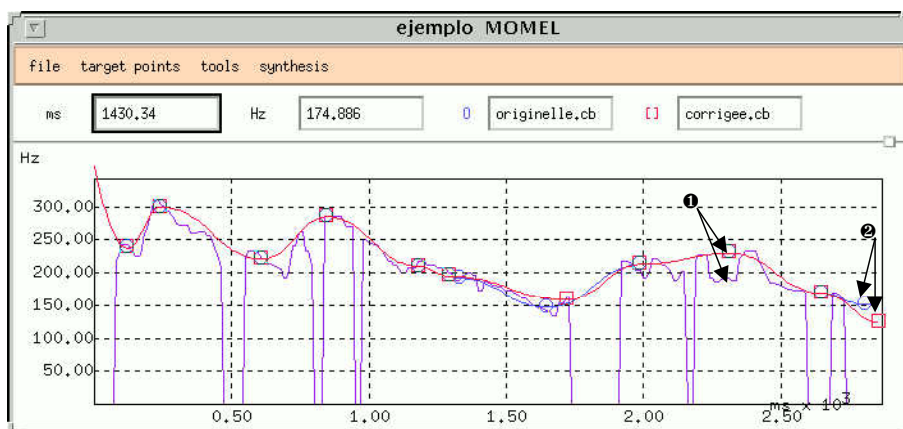


Figura 4. Ejemplo de extracción y modelización automática de la curva melódica y de su posterior corrección manual.

Así, en el ejemplo que nos ocupa, y que ilustramos en la Figura 4, la curva estilizada obtenida automáticamente ya no contempla las variaciones microprosódicas de la curva obtenida en el nivel físico, como se puede observar a partir de la diferencia que existe entre los valores de frecuencia fundamental correspondientes a la [d] de “diabétique” entre las dos curvas (véase Figura 4, ❶).

Por otra parte, hemos elegido un ejemplo en el que se ha tenido que corregir manualmente la curva estilizada obtenida automáticamente para conseguir una modelización perceptivamente equivalente a la original. Se pueden observar los efectos de dicha corrección al final del enunciado, ya que ha sido preciso bajar el valor de F0 del último punto de inflexión (véase Figura 4, ❷) para que se perciba como una declarativa completa.

5.2.3. Nivel fonológico superficial

Este nivel se define a partir de la hipótesis de que la prosodia de lenguas diferentes se puede describir mediante una misma serie de símbolos; se trata, pues, del nivel de generalización donde se pueden describir los patrones entonativos de las distintas lenguas.

Los puntos de inflexión obtenidos con el algoritmo MOMEL son codificados mediante el sistema de anotación INTSINT (Hirst, Di Cristo y Espesser, 2000). Cada uno de los puntos de la curva estilizada es codificado como un **tono absoluto** —definido globalmente en referencia al rango frecuencial del locutor— o como un **tono relativo** —definido localmente en relación a los puntos de inflexión anterior y posterior. Entre los tonos relativos hay que distinguir los **iterativos** —definidos como puntos de inflexión intermedios en secuencias ascendentes o descendentes— y los **no iterativos** —definidos como puntos de inflexión que no corresponden a tonos absolutos y que no constituyen etapas intermedias en secuencias ascendentes o descendentes.

Los tonos absolutos son los siguientes:

- T (*Top*): altura tonal máxima del locutor (↑↑)
- B (*Bottom*): altura tonal mínima del locutor (↓↓)
- M (*Mid*): etiqueta inicial correspondiente a un valor medio del locutor (⇒)

VERSIÓN BORRADOR PROVISIONAL

Los tonos relativos son estos:

- no iterativos:
 - H (*Higher*): punto más alto que los dos puntos adyacentes (\uparrow)
 - L (*Lower*): punto más bajo que los dos puntos adyacentes (\downarrow)
 - S (*Same*): punto igual que su precedente (\rightarrow)
- iterativos:
 - U (*Up – Upstepped*): punto en una secuencia ascendente ($<$)
 - D (*Down – Downstepped*): punto en una secuencia descendente ($>$)

Su ejemplificación es la que se observa en la Figura 5¹⁴:

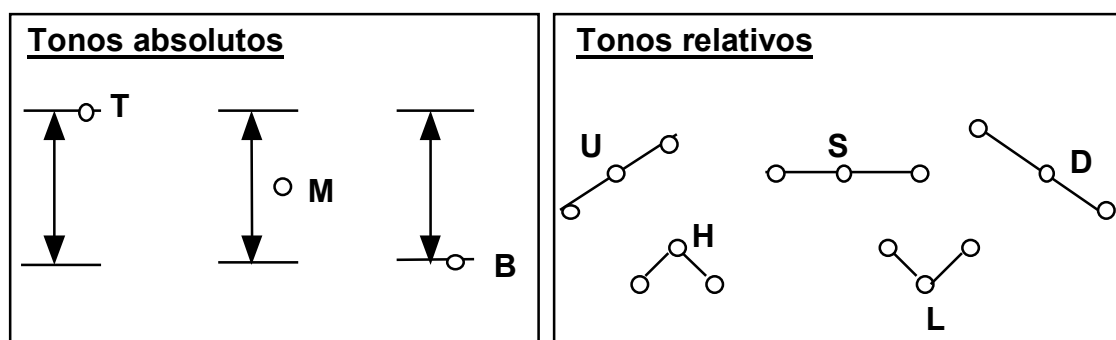


Figura 5. Etiquetaje de tonos con el sistema INTSINT.

Las etapas que sigue el algoritmo de codificación de la curva de F0 mediante el sistema INTSINT son las siguientes:

- a) Los puntos con el valor máximo y mínimo de frecuencia son codificados como T y B respectivamente.
- b) El primer punto de la curva es codificado como M a no ser que ya haya sido codificado como T o B.
- c) Los otros puntos son codificados como relativos. Un tono que no supera un límite determinado de frecuencia respecto al punto anterior se etiqueta como S; en los otros casos, las etiquetas H, L, U y D se adjudican a los puntos teniendo en cuenta la relación que mantienen con los puntos anterior y posterior. También pueden ser codificados como T, M o B si la frecuencia de estos puntos es muy similar a la de los puntos máximo y mínimo (ver etapas a) y b)).
- d) El paso siguiente consiste en calcular los valores estadísticos de cada una de las categorías de puntos. En el caso de los tonos absolutos, se calcula la media, mientras que en los relativos se calcula una recta de regresión partiendo del punto inmediatamente anterior.
- e) Un punto situado tras una pausa de más de 250 milisegundos¹⁵ puede ser recodificado como M. Puntos que han sido codificados como T, U, B o D pueden ser recodificados si esto comporta una mejora del modelo estadístico.

¹⁴ Para una definición más detallada y acorde con el algoritmo, ver Campione, Hirst y Véronis (2000).

¹⁵ Para una discusión más detallada del tratamiento de las pausas, véase el apartado 3.25.4.

VERSIÓN BORRADOR PROVISIONAL

- f) Las etapas d) y e) se repiten hasta que no quede ningún punto que pueda ser recodificado.

Podemos visualizar este proceso a partir del ejemplo de la frase catalana *És imprescindible que la tingui demà, perquè sóc diabètic*. Así, una vez terminadas las fases correspondientes a los niveles físico y fonético ya mencionadas, un algoritmo codifica automáticamente los puntos de inflexión seleccionados mediante el alfabeto INTSINT y, en el caso de nuestro ejemplo, se obtiene el resultado siguiente (véase Figura 6), que contempla la señal acústica alineada con:

- a) un etiquetaje (opcional) de las unidades lingüísticas (aquí, las palabras del enunciado) y el etiquetaje INTSINT que se corresponde con la codificación automática de los puntos de inflexión, y
b) las curvas melódicas del nivel físico y del fonético.



Figura 6. Ejemplo del proceso de extracción y modelización de la curva melódica, alineada con el etiquetaje automático mediante INTSINT.

5.2.4. Nivel fonológico profundo

VERSIÓN BORRADOR PROVISIONAL

En este nivel, una serie de selecciones lingüísticamente significativas interactúan con un conjunto de parámetros prosódicos abstractos específicos de cada lengua para generar los patrones entonativos concretos.

Las reglas que determinan el patrón entonativo de un enunciado según este modelo son las siguientes:

- a) división del enunciado en unidades entonativas,
- b) división de las unidades entonativas en unidades tonales,
- c) asignación de los patrones entonativos a las unidades entonativas,
- d) asignación de los patrones entonativos a las unidades tonales,
- e) aplicación, si es preciso, de la regla de reajuste del **escalonamiento descendente** (*downstep*) según la cual se baja un tono alto (H) cuando viene precedido por un tono alto seguido de un tono bajo (HL) de tal manera que existe un descenso progresivo de la altura de los picos tonales,
- f) aplicación, si necesario, de otras reglas de reajuste.

Se puede obtener la secuencia de símbolos INTSINT que dan cuenta de la estructuración prosódica del enunciado por un proceso ascendente como el que acabamos de ilustrar, pero también por un proceso descendente de índole propiamente fonológica. Así, si aplicamos las reglas que determinan el patrón entonativo de un enunciado según el modelo de Aix-en-Provence al enunciado catalán de nuestro ejemplo, se obtiene el razonamiento siguiente:

- 1) Se establece el árbol fonológico y la rejilla métrica asociada al enunciado. En este nivel puede ser necesario tomar en cuenta todos aquellos aspectos que tienen que ver con la acentuación enfática o distintiva, la desacentuación en presencia de *clash* acentuales¹⁶, etc. Así, se divide el enunciado en unidades entonativas (UE), unidades rítmicas (UR) y unidades tonales (UT) de la manera siguiente:
 - a) [És imprescindible que la tingui demà]_{UE} [perquè sóc diabètic]_{UE}
 - b) [{És imprescindible}_{UR} {que la tingui demà}_{UR}]_{UE} [{perquè sóc}_{UR} {diabètic}_{UR}]_{UE}
 - c) [{(És)_{UT} (im pres cin di) _{UT} ble}_{UR} {(que la tin) _{UT} (gui de mà) _{UT} }_{UR}]_{UE} [{(per què)_{UT} (sóc)_{UT} }_{UR} {(di a bè) _{UT} tic}_{UR}]_{UE}

3	(+ . . . + . . . + . . +) (. . + . . + .)
2	(+ . . . + .) (. . + . . +) (. + +) (. . + .)
1	(+) (+ + + +) + (+ + +) (+ + +) (+ +) (+) (+ + +) +
S	És im Pre: cin di ble que la tin gui de mà per què sóc di a bè tic

Figura 7. Árbol fonológico y rejilla métrica del enunciado catalán *És imprescindible que la tingui demà, perquè sóc diabètic*. (S) es el nivel de las sílabas, (1) el de las unidades tonales, (2) el de las unidades rítmicas, y (3) el de las unidades entonativas. [+ : sílaba proeminente, . : sílaba no proeminente].

- 2) Se atribuyen los tonos que se postulan a priori a las unidades tonales (LH) y entonativas (en este caso, LL)¹⁷, según la lengua considerada y la estructura del

¹⁶ Se habla de *clash* acentual o de choque acentual cuando dos sílabas sucesivas aparecen acentuadas.

¹⁷ En el apartado 5.3 se describen los patrones básicos de las distintas modalidades oracionales.

VERSIÓN BORRADOR PROVISIONAL

enunciado. Así, en el caso de nuestro ejemplo, se obtiene el patrón entonativo siguiente¹⁸:

[L (LH) (LH) (LH) (LH) L] [L (LH) (LH) (LH) L]

- 3) Se aplican las reglas de reajuste y linearización siguientes¹⁹:
- Conforme a los principios de simplificación, los tonos L iniciales de la representación anterior fusionan en un tono medio M característico del registro de referencia del locutor.
 - El tono L final se realiza como un tono bajo absoluto B que corresponde al valor límite inferior de la dinámica tonal del locutor.
 - Según la regla de reajuste de escalonamiento descendente que se aplica a un tono H precedido de una secuencia de tonos H y L en la misma unidad entonativa, éste es bajado y se realiza como tono D por la presencia del tono L. Posteriormente, desaparece el tono bajo que ha provocado el escalonamiento descendente. En el caso que nos ocupa, esta regla se aplica en tres ocasiones.
 - Aplicando la regla de escalonamiento ascendente (*upstep*), el tono L precedido de una secuencia de tonos L y H en la misma unidad entonativa, este último no se realiza fonéticamente pero provoca un realce del tono L siguiente, que se realiza como un tono U.
 - Los segmentos tonales de las distintas unidades prosódicas se sitúan en líneas diferentes, por lo cual es necesario aplicar una regla de linearización según la cual se copian los tonos de las unidades de rango superior en la estructura prosódica en las unidades tonales adecuadas. Se obtiene así una secuencia fonológica pronunciable.
 - Según la regla de simplificación tonal, se reducen las secuencias de segmentos tonales idénticos en la misma unidad tonal en un único segmento tonal a fin de evitar la colisión de dos tonos idénticos en una única unidad.
 - El tono H inicial se realiza como un tono absoluto T correspondiente al límite superior de la dinámica tonal del locutor, ya que corresponde al máximo frecuencial de la oración.

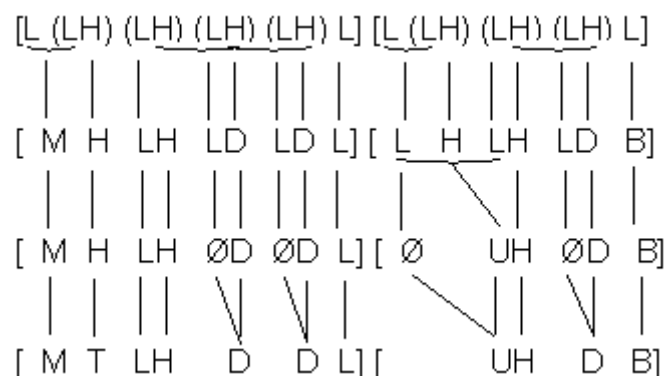


Figura 8. Ejemplo de aplicación de las reglas de reajuste y linearización para el enunciado *És imprescindible que la tingui demà, perquè sóc diabètic*.

¹⁸ En esta fase, no se toman en consideración las unidades rítmicas, ya que, como se ha indicado más arriba, éstas no inciden sobre la configuración tonal, sino sobre otros parámetros.

¹⁹ Se hallarán los fundamentos teóricos de dichas reglas y sus modalidades de aplicación en Hirst y Di Cristo (1984).

Se obtiene así una secuencia de tonos linearizada apta para ser pronunciada de la manera siguiente, que corresponde a la secuencia [MTLHDDLHDB] obtenida automáticamente por el sistema. Cabe señalar, sin embargo, que la secuencia obtenida podría variar si se incorporara en el modelo alguna técnica de normalización que permita escalar los datos entre distintas unidades entonativas (véase apartado 5.4).

5.3. Aplicaciones del modelo

Como se habrá podido deducir de los apartados dedicados a la descripción del modelo, más allá de la elaboración de una teoría que contemple tanto aspectos considerados universales como reglas que permitan llevar a cabo un comparatismo interlingüístico y, por lo tanto, dar cuenta de las especificidades de la estructuración prosódica de cada una de las distintas lenguas, se trata de realizar descripciones pormenorizadas de grandes *corpora* para poder validar las distintas hipótesis que fundamentan el modelo y, precisar, si tal es el caso, los parámetros específicos de la lengua objeto de estudio.

La versión actual del modelo es muy reciente, desarrollada a partir de 1995, y son muchos los trabajos parciales o en preparación que contribuirán a aplicarlo a distintas lenguas²⁰. Por el momento, el francés y el inglés constituyen los sistemas descritos con más detalle, aunque distintas tesis doctorales y artículos dan cuenta de descripciones generales o más particulares de otros idiomas. Prueba de ello es la publicación editada por Hirst y Di Cristo en 1998 en la que se presentan, mayoritariamente según los principios de dicho modelo, las principales características prosódicas de 20 lenguas, entre las cuales el español. Con todo, es necesario desarrollar todavía investigaciones más sistemáticas para poder precisar mejor las hipótesis de distintos sistemas lingüísticos.

Para el español, cabe citar los estudios de Mora (1996), Le Besnerais (1995), Alcoba, Le Besnerais y Murillo (1993), Alcoba y Murillo (1998). En este apartado aplicaremos el modelo al catalán, lengua para la cual todavía existen menos estudios (Estruch, 1999; Riera, 2001), pero que comparte con el castellano una serie de características básicas, como el hecho de poseer, contrariamente al francés, un acento léxico libre, esto es, que puede recaer en distintas sílabas de una unidad léxica; este acento, además, es pertinente léxicamente, en la medida en que permite oponer palabras como *làmina*, *lamina* y *laminà*²¹. Por otra parte, en ambos casos, se genera una rejilla métrica en función de los acentos de tal suerte que los grupos tonales estén compuestos de una sílaba acentuada y las sílabas átonas adyacentes, y que las sílabas átonas posteriores al último acento se consideren como extramétricas.

En la aplicación que proponemos, nos limitamos a ilustrar algunos fenómenos generales que caracterizan la estructuración prosódica del catalán y que, en gran medida, podrían aplicarse asimismo al castellano: los patrones básicos de los enunciados declarativos no enfáticos en función del número de grupos tonales que contienen y de la posición de la sílaba acentuada en la palabra, y las distintas modalidades oracionales.

²⁰ Se puede consultar la revista *Travaux de l'Institut de Phonétique d'Aix* y la página <http://www.lpl.univ-aix.fr>, que se actualiza regularmente con las nuevas publicaciones y trabajos que se realizan en el marco del Laboratoire Langage et Parole de la Université d'Aix-en-Provence.

²¹ En español, respectivamente, *lámina*, *lamina* y *laminó*.

Baqué, Lorraine y Estruch, Mónica (2003):
"Modelo de Aix-en-Provence", in Prieto, Pilar (ed.): *Teorías de la entonación*. Barcelona (España), págs. 123-153.
VERSIÓN BORRADOR PROVISIONAL

5.3.1. Patrones básicos de frases declarativas no enfáticas

A) Frases declarativas de un grupo tonal

El patrón típico de las frases declarativas no enfáticas en catalán para los casos que contienen un único grupo tonal se caracteriza por los siguientes rasgos: la presencia de un punto inicial M; un único pico (punto T) que coincide aproximadamente con la sílaba tónica y un punto B al final del enunciado.

Veamos tres ejemplos, cada uno con el acento en una sílaba diferente, en los que observamos que el pico se desplaza en función de la situación de la sílaba tónica:

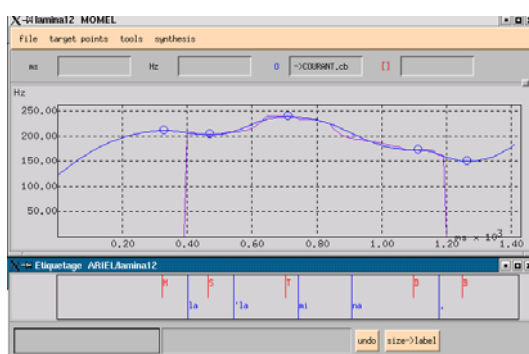


Figura 9. Patrón melódico y transcripción tonal correspondiente al enunciado declarativo *La làmina*.²²



Figura 10. Patrón melódico y transcripción tonal correspondiente al enunciado declarativo *La lamina*.²³

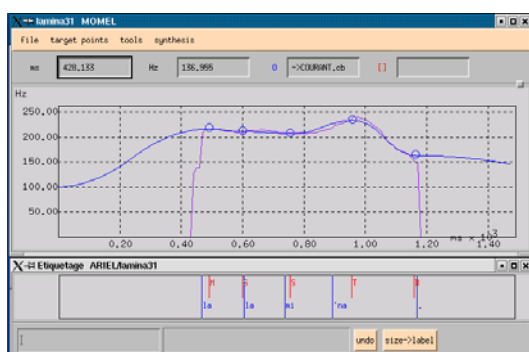


Figura 11. Patrón melódico y transcripción tonal correspondiente al enunciado declarativo *La laminà*.²⁴

²² En castellano, *La lámina*.

²³ En castellano, *La lamina*.

²⁴ En castellano, *La laminó*.

VERSIÓN BORRADOR PROVISIONAL

Como se puede observar en las Figuras 9, 10 y 11, el patrón melódico global de las frases declarativas de un grupo tonal se caracteriza por una rama ascendente y una rama descendente, independientemente de la posición de la sílaba acentuada. Con todo, teniendo en cuenta el hecho de que el modelo intenta estilizar la curva melódica de tal modo que sea perceptivamente equivalente a la original, y que la transcripción mediante el alfabeto INTSINT viene determinada por dicha modelización, es obvio que aparecen puntos intermedios que corresponden a sílabas átonas y que varían según la distancia que existe entre la sílaba acentuada (punto T) y los puntos inicial (M) y final (B). A ello se deben las diferencias de secuencias de símbolos que se corresponden con las frases:

- La làmina.* [MSTDB]
- La lamina.* [MSTDB]
- La laminà.* [MSSTB]

B) Frases declarativas de más de un grupo tonal

En el caso de que el enunciado tenga dos sílabas acentuadas (esto es, dos grupos tonales), el patrón entonativo se caracteriza por: un punto inicial M; un punto T que coincide con la primera sílaba tónica; un segundo acento (correspondiente a la segunda sílaba tónica) antes de la pausa final que mantiene el nivel de F0 a una altura media con escalonamiento descendente (D) que termina en un nivel B. Así, por ejemplo, en la frase *La mare de la Maria*²⁵ obtenemos la Figura 12.

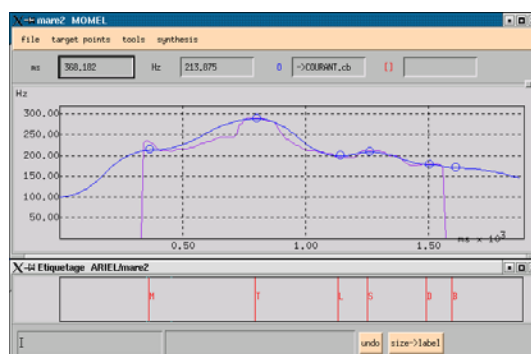


Figura 12. Patrón melódico y transcripción tonal correspondiente al enunciado *La mare de la Maria*.

Como se puede observar en la Figura 12, la secuencia tonal resultante es [MTLSDB]. Esta configuración tonal corresponde a la rejilla métrica del enunciado, formado por un único grupo entonativo a los límites del cual se atribuyen dos tonos L por tratarse de una frase declarativa²⁶. Por otra parte, a cada grupo tonal se atribuyen a priori los tonos (LH), dónde la H corresponde a la sílaba acentuada.

Así el enunciado corresponde a la estructura tonal abstracta siguiente:

[(La **ma**)(re de la **Mari**)a]
[L(L H) (L H)L]

Según las reglas de reajuste y linearización ya mencionadas en el apartado 5.2.4:

- Los tonos L iniciales de la representación fusionan en un tono medio M característico del registro de referencia del locutor (principio de simplificación).

²⁵ En castellano, *La madre de María*.

²⁶ En el apartado siguiente, se describen los patrones básicos de las distintas modalidades oracionales.

VERSIÓN BORRADOR PROVISIONAL

- El tono L final se realiza como un tono bajo absoluto B que corresponde al valor límite inferior de la dinámica tonal del locutor.
- El segundo tono H es bajado y se realiza como tono D por la presencia del tono L, ya que viene precedido de una secuencia de tonos H y L en la misma unidad entonativa. Posteriormente, desaparece el tono bajo que ha provocado el descenso (regla del escalonamiento descendente).
- El tono H inicial se realiza como un tono absoluto T correspondiente al límite superior de la dinámica tonal del locutor.

Aplicando estas reglas, se obtiene la secuencia tonal [MTDB]. La aparición de los tonos LS entre T y D se debe a la existencia de 4 sílabas átonas entre los acentos. En efecto, entonces aparece un acento secundario de índole rítmica, que constituiría un tercer grupo tonal de menor importancia, al cual se atribuiría igualmente los tonos (LH), y, si aplicamos de nuevo las reglas de reajuste y linearización, obtenemos la secuencia etiquetada que observamos en la figura anterior, esto es, [MTLSDB].

Finalmente, en el caso de que el enunciado declarativo tenga tres grupos tonales, correspondientes a tres sílabas acentuadas, el patrón se caracterizará por los siguientes rasgos: un punto M inicial; un punto máximo T coincidente con la primera sílaba tónica; un punto H que coincide con la segunda sílaba tónica; y un progresivo escalonamiento descendente (D) al final hasta llegar al punto mínimo (B) prepausal, por la regla del escalonamiento descendente ya mencionada en el ejemplo anterior. Un ejemplo de ello es la frase declarativa *La mare de l'àvia de la Maria*²⁷, que corresponde a la secuencia INTSINT siguiente. La **mare** de l'**à**via de la **Maria**
[M T L H D B]

Como se puede observar de los ejemplos anteriores, lo que caracteriza cada grupo tonal es la presencia de un pico de F0, más alto el primero y progresivamente inferiores los demás, debido a la regla del escalonamiento.

5.3.2. Patrones básicos de las distintas modalidades oracionales

Las distintas modalidades oracionales presentan una serie de características que permiten diferenciarlas entre ellas. En este apartado veremos una misma frase producida con entonaciones diferentes y comprobaremos similitudes y diferencias para poder caracterizar cada una de las modalidades oracionales.

²⁷ En castellano, *La madre de la abuela de María*.

A) Declarativa

Las características del patrón de una frase declarativa en su forma superficial son las siguientes: un punto M inicial, un punto B que marca el fin del enunciado y una serie variable de picos que corresponden a las sílabas acentuadas. Dichos picos aparecen, si no existe ninguna marca enfática, en una secuencia descendente, de tal suerte que el último suele manifestarse no por un pico real, sino por un tono D, debido a las reglas de reajuste que ya hemos mencionado. Esto se puede observar en la figura correspondiente al enunciado *Voldria reservar una taula*²⁸.

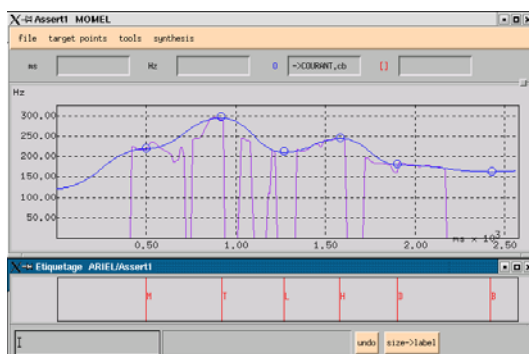


Figura 13. Patrón melódico y transcripción tonal correspondiente al enunciado declarativo *Voldria reservar una taula*.

Así, en el caso del enunciado declarativo de la Figura 13, al estar formado por tres grupos tonales (véase 5.3.1), obtenemos la siguiente transcripción: [MTLHDB].

B) Interrogativa total

Una oración interrogativa total de tres grupos tonales se caracteriza por los siguientes rasgos: un punto inicial M; un pico H (le correspondería un punto T, pero al no poder aparecer dos puntos T en una misma oración, éste, más bajo, se codifica como H); un escalonamiento descendente (DB) para recuperar el nivel más alto de F0 al final (T). Un ejemplo de ello se observa en la Figura 14.

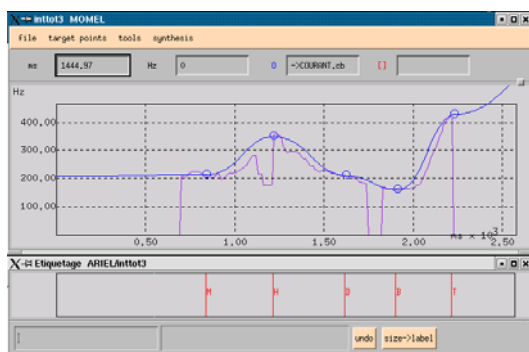


Figura 14. Patrón melódico y transcripción tonal correspondiente al enunciado interrogativo total *Voldria reservar una taula?*²⁹

²⁸ En castellano, *Quisiera reservar una mesa*.

²⁹ En castellano, *¿Quisiera reservar una mesa?*

VERSIÓN BORRADOR PROVISIONAL

Se obtiene la secuencia de tonos INTSINT [MHDBT] de la manera siguiente:

- a) se atribuye a priori al enunciado, por tratarse de una interrogativa total, la secuencia [LH];
- b) se segmenta el enunciado en tantos grupos tonales como sílabas acentuadas, en este caso tres: (vol-DRI)(a-re-ser-VAR-U)(na TAU)la; la última sílaba, al ser átona, es extramétrica;
- c) se atribuye a priori a cada grupo tonal la secuencia (LH);
- d) se aplican las reglas de reajuste y linearización, según las cuales:
 - Los dos tonos L iniciales fusionan en un tono M.
 - Los dos tonos H finales fusionan en un único punto H, que se realiza como un tono alto absoluto T que corresponde al valor límite superior de la dinámica tonal del locutor.
 - Según la regla de reajuste de escalonamiento descendente que se aplica a un tono H precedido de una secuencia de tonos H y L en la misma unidad entonativa, éste es bajado y se realiza como tono D por la presencia del tono L. Posteriormente, desaparece el tono bajo que ha provocado el escalonamiento descendente.
 - Se recodifica el tono L que queda en B por tratarse del punto más bajo de la curva que corresponde, por tanto, al límite inferior de la dinámica tonal del locutor.

Estas fases se pueden esquematizar de la manera siguiente:

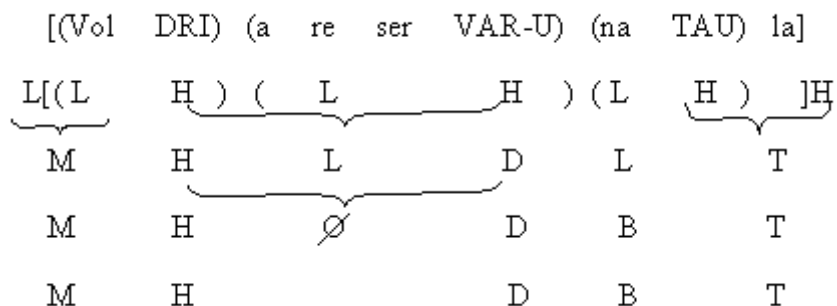


Figura 15. Aplicación de las reglas de reajuste y linearización para el enunciado *Voldria reservar una taula?*

C) Interrogativa parcial

Las características de una interrogativa parcial son las siguientes: un punto M inicial; un pico (T) que coincide con el elemento interrogativo (“qui”); y un escalonamiento progresivo descendente que culmina en el punto de F0 más bajo, B. Esto se puede observar en la figura siguiente:

VERSIÓN BORRADOR PROVISIONAL

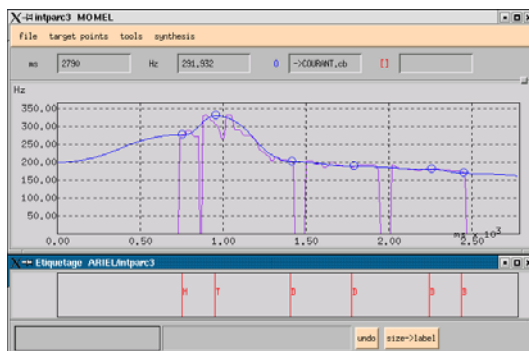


Figura 16. Patrón melódico y transcripción tonal correspondiente al enunciado interrogativo parcial *Qui voldria reservar una taula?*³⁰

En este caso, la oración interrogativa parcial tiene cuatro grupos tonales, ya que el pronombre interrogativo constituye una palabra acentuada, por lo que se obtiene la secuencia de símbolos [MTDDDB].

D) Imperativa

Finalmente, las frases imperativas como la que muestra la Figura 17, se caracterizan por un punto M inicial; un punto máximo T que coincide generalmente con la primera sílaba tónica; y un escalonamiento descendente que termina en el punto B final.

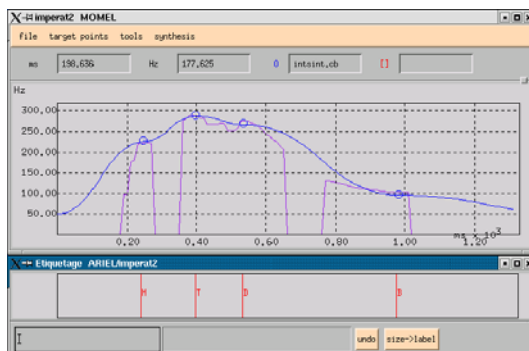


Figura 17. Patrón melódico y transcripción tonal correspondiente al enunciado imperativo *Reserva una taula!*³¹

Las diferencias que se observan entre esta modalidad y las anteriores se deben a que se atribuye al enunciado el patrón [HL]. Así, aplicando de manera reiterada las reglas de reajuste que hemos ilustrado en los casos anteriores, se obtiene la secuencia [MTDB], que corresponde a un enunciado imperativo como el que ilustramos, que contiene dos grupos tonales: (*re-SER*) (*va-u-na-TAU*)*la*.

³⁰ En castellano, *¿Quién quisiera reservar una mesa?*

³¹ En castellano, *¡Reserva una mesa!*

5.4. Discusión

El modelo de Aix-en-Provence, como hemos visto, pretende proponer un modelo que permita no sólo derivar una salida acústica a partir de una entrada simbólica, sino también llevar a cabo el planteamiento inverso que consiste en obtener una representación simbólica a partir de la curva de F0.

Sin embargo, en su estado actual, y como lo reconocen sus mismos autores, el modelo de Aix-en-Provence no constituye todavía un modelo fonético completo de la entonación, y quedan algunos aspectos que no han sido explorados o desarrollados suficientemente. Mencionaremos, particularmente, los aspectos siguientes:

- la declinación.
- los reajustes frecuenciales y las pausas.

A) Declinación

La **declinación** se define como la tendencia universal de la frecuencia del fundamental a declinar a lo largo de un enunciado³². Este concepto fue descrito por Pike, ya en el año 1945 y por Bolinger (1964) a partir de datos perceptivos. Cohen y t'Hart (1965) usaron el término en estudios acústicos. Uno de los resúmenes más completos de los estudios realizados sobre la declinación es el que se encuentra en Ladd (1984).

Existen dos aproximaciones al estudio de la declinación: por un lado, el punto de vista puramente estadístico, según el cual puede afirmarse que existen más tipos de contornos de F0 que declinan que no contornos que no lo hagan, y que los primeros son mucho más utilizados que los segundos. Por otro lado, existe el punto de vista del marco de referencia: la declinación se trata como una modificación sistemática, a lo largo de la producción del enunciado, del marco de referencia fonético en que se realizan los elementos fonológicos. Esto se puede entender como un telón de fondo sobre el cual se sitúan los fenómenos locales de F0, que cambia gradualmente. Estos cambios serían consecuencia directa de los mecanismos fisiológicos implicados en la producción del habla: el descenso del F0 es debido al descenso de la presión subglotal que acompaña la salida del aire de los pulmones durante el habla³³.

Se han hecho multitud de estudios sobre la declinación, en lenguas diferentes y también tomando diferentes niveles o ámbitos de actuación. Citaremos, entre otros, el de Nicolas (1995) y Garrido (1993, 1996) que analizan la declinación en párrafos; Lehiste (1975) que analiza el nivel de la oración; Grønnum Thorsen (1985, 1986) describe la declinación en secuencias semánticamente coherentes; Cooper y Sorensen (1981), McAllister (1971) y O'Shaughnessy (1976) que no definen un ámbito de actuación, sino que propugnan la hipótesis de que el hablante utiliza un mecanismo de preplanificación que tiene en cuenta la longitud aproximada del enunciado.

En la definición del modelo de Aix-en-Provence que estamos analizando, se afirma que no se tiene en cuenta el fenómeno de la declinación. Campione, Hirst y Véronis (2000) afirman que "hay que tener en cuenta que todavía no hay un componente de *downdrift* o

³² Otros términos ingleses que se usan a menudo de manera equivalente son *downdrift* o *downtrend*.

³³ Esta hipótesis es la defendida por autores como Lieberman (1967), Bruce (1977), Thorsen (1980), Cooper y Sorensen (1981) y Fujisaki (1983), entre otros.

VERSIÓN BORRADOR PROVISIONAL

de declinación en el modelo, el típico escalonamiento descendente de una secuencia de puntos H y L sucesivos observados en el habla natural es una consecuencia automática de aplicar iterativamente los mismos coeficientes de regresión". Sin embargo, en algunos estudios realizados utilizando este modelo (Nicolas, 1995; Estruch, 1999) se ha visto que sí existe una declinación gradual de los enunciados.

Como puede observarse en la Figura 18, correspondiente a nuestra oración de ejemplo, los picos cada vez presentan una altura menor a lo largo de la oración, lo mismo que los puntos situados en valles (la frecuencia del fundamental es cada vez más baja).

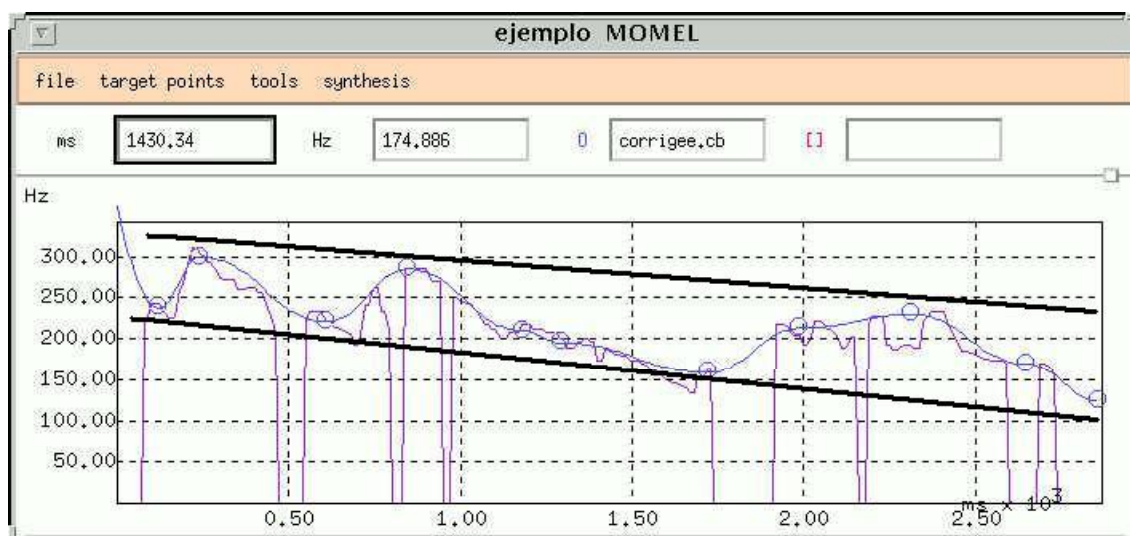


Figura 18. Curva modelizada de la oración *És imprescindible que la tingui demà, perquè sóc diabètic.*, donde puede observarse la declinación general de la oración, tanto de la línea superior como de la inferior.

B) Reajustes frecuenciales y pausas

Íntimamente ligado al aspecto de la declinación está el del **reajuste frecuencial** de la curva. Este fenómeno se define como la recuperación de la curva de F0 en una nueva unidad entonativa, que comienza en un punto frecuencial más alto que aquél donde acabó. Hay varios autores que han estudiado este fenómeno: Cooper y Sorensen (1977) en el caso del inglés americano, Collier (1985) en el holandés y Garrido (1996) en el español.

Sorensen y Cooper (1980) distinguen dos tipos de reajuste tonal: el total o completo y el parcial. En el primer caso, un **reajuste total** se define porque el inicio de la línea superior o inferior se sitúa al mismo nivel o por encima del inicio de la línea precedente; por otro lado, el **reajuste parcial** se da cuando el inicio de la línea superior o inferior se sitúa por debajo del inicio de la línea precedente.

En la figura siguiente, correspondiente a nuestra oración de ejemplo, podemos observar un caso de reajuste parcial.

VERSIÓN BORRADOR PROVISIONAL

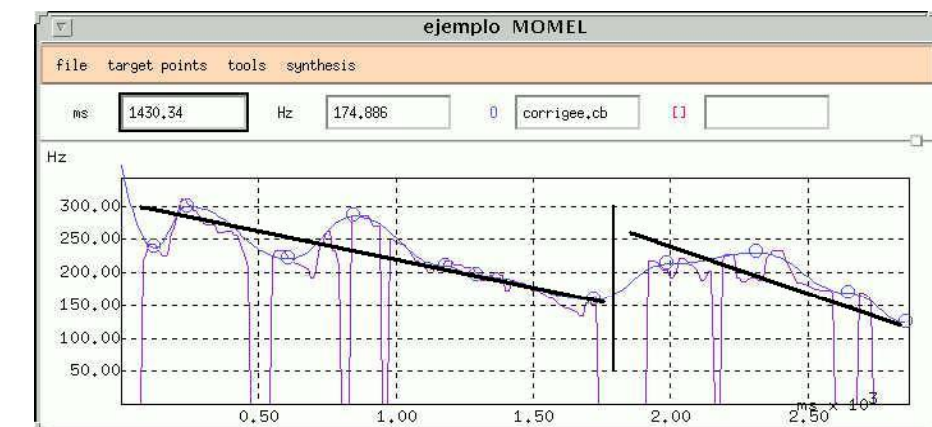


Figura 19. Ejemplo de reajuste parcial (la línea de declinación después de la pausa —marcada con una línea vertical— empieza en un nivel superior al del final de la línea de declinación anterior a la pausa).

Los reajustes frecuenciales totales se dan bajo tres condiciones, según Garrido (1993):

- es necesaria la presencia de una pausa o de una frontera sintáctica que explique la presencia de un reajuste (marcada con una línea vertical en la Figura 19),
- el valor del último punto máximo antes de la posible frontera tiene que ser más bajo que el valor del primer punto máximo después de la frontera,
- el primer valor mínimo de F0 antes de la frontera también tiene que ser más bajo que el primer valor mínimo después de la frontera.

En este caso, la pausa de 421 ms. propicia la aparición de un reajuste parcial. El valor de F0 del último punto antes de la frontera, situado a 150 Hz. es inferior al valor del primer punto después de la frontera, que se sitúa a 214 Hz.

Finalmente, otro aspecto interrelacionado con los dos anteriores es la relación existente entre las pausas y los reajustes frecuenciales. Las pausas son definidas por Drommel (1980) como “intervalos de silencio de una determinada duración que reúnen unas condiciones determinadas (principalmente, una reducción de la energía a un nivel cero durante un tiempo)”. En las indicaciones del programa MES correspondientes al modelo que estamos analizando, se especifica como valor de referencia para la duración de una pausa unos 250 milisegundos aproximadamente.

La aparición de las pausas es un tema que plantea ciertos problemas en la aplicación del modelo. Mencionaremos algunos de ellos en el apartado siguiente, aunque pueden encontrarse descripciones más detalladas de estos desajustes en Llisterri (1996) o Estruch (1999).

C) Limitaciones y propuestas futuras del modelo

Teniendo en cuenta que en dicho modelo se contempla la entonación como un fenómeno que se manifiesta en las dimensiones de variaciones de frecuencia fundamental, de intensidad y duración, queda por incorporar una representación de las características de los dos últimos parámetros que todavía no aparecen de manera explícita en la representación fonológica de la entonación actualmente propuesta, aunque la inclusión de la unidad rítmica en el modelo y primeros estudios sobre la posible codificación de la duración³⁴ y la intensidad permiten anunciar próximos desarrollos en este aspecto.

Por otra parte, en lo que al nivel fonético se refiere, queda por resolver una dificultad importante que consiste en escalar de manera relativa los puntos de inflexión, en particular entre producciones de distintos locutores. La problemática de la normalización de los datos acústicos tiene una relevancia capital si se pretende obtener un modelo independiente del locutor, tanto más cuanto que las características prosódicas pertinentes suelen corresponder a valores relativos más que a valores absolutos. Con todo, por el momento no se ha desarrollado en el modelo un método que permita dicha normalización, y los autores apuntan que se podrían utilizar técnicas basadas en condicionantes universales de producción y percepción como las que proponen Hermes y van Gestel (1991), Campbell (1992) o Terken y Hermes (2000).

En el nivel de la representación fonológica superficial aparece también un problema de normalización de datos, en particular en lo que se refiere a escalar los tonos absolutos (T, M y B) en producciones más extensas que un enunciado o un párrafo, ya que, como se habrá observado, la atribución de tonos absolutos depende actualmente de manera directa del segmento fónico analizado. Si se pretende obtener un modelo entonativo general, es imprescindible hallar un método que permita dar cuenta, por una parte, del registro del locutor, y, por otra parte, de la estructuración de unidades discursivas más amplias. En efecto, como lo demuestran los estudios de Nicolas (1995) o Nicolas y Hirst (1995), en textos continuos, los resultados son mucho más concluyentes si se incluye alguna técnica que permita escalar los datos obtenidos, y así dar cuenta del hecho que, en francés por ejemplo, el primer tono T de una unidad entonativa tiende a ser significativamente más alto que la media en la primera unidad entonativa y significativamente más bajo en la última unidad entonativa de un párrafo ("paratono"). Del mismo modo, los tonos B finales suelen ser significativamente más bajos al final de unidades amplias de este tipo.

Aunque no se haya implementado todavía en el modelo una técnica concreta que permita resolver dicha dificultad, los autores, que asumen la necesidad de escalar de algún modo las diferencias que existen entre las distintas unidades entonativas, proponen dos posibles soluciones. La primera consiste en usar los símbolos INTSINT para codificar no únicamente los puntos de inflexión sino el rango y el registro de las unidades entonativas tal como lo recomienda el grupo de trabajo sobre transcripción suprasegmental del IPA (véase Bruce, 1989). Así, en el ejemplo que nos ha ocupado en el apartado 5.2, se obtendría la transcripción siguiente:

És imprescindible que la tingui demà, perquè sóc diabètic.

³⁴ [Cf. Véase, por ejemplo, Hirst \(1999\).](#)

VERSIÓN BORRADOR PROVISIONAL

(T) [M T L H D D B] (D) [M T D B]

donde las unidades entonativas vienen delimitadas por corchetes y los símbolos que aparecen en ellas corresponden a los puntos de inflexión, mientras que los símbolos entre paréntesis definen el rango tonal de la unidad entonativa siguiente.

Otra solución propuesta por Hirst, Di Cristo y Espesser³⁵ para resolver la misma problemática consistiría en considerar el rango tonal T y D respectivamente en nuestro ejemplo como valores por defecto de tal suerte que la primera unidad entonativa tendría asociado un rango T y la segunda un rango D, con lo cual se obtendría una transcripción simplificada [MTLHDDDB] [MTDB].

Sin embargo, todavía no se ha desarrollado en el marco del modelo de Aix-en-Provence un algoritmo que permita obtener las representaciones que acabamos de proponer. Cabe señalar que dividir un segmento fónico en unidades entonativas de manera automática sigue sin ser tarea fácil³⁶, puesto que hay que definir una serie de parámetros tales como los valores por encima de los cuales se puede considerar que existe una reestructuración tonal o la duración de un silencio para que pueda ser considerado una pausa. En lo que respecta a este parámetro, se trata de un valor que varía significativamente de un locutor a otro o de un estilo de habla a otro, y que además constituye actualmente una de las causas principales de error del algoritmo MOMEL, que tiende a alisar la curva melódica si obtiene una variación brusca de F0 cuando no lo atribuye a la presencia de una pausa. Actualmente, la única posibilidad para adaptar el algoritmo a la duración media de las pausas es modificar manualmente el parámetro en función de los valores observados, pero los autores abogan por implementar en una fase posterior un algoritmo que permita a la vez una segmentación en unidades entonativas y una técnica para escalar los tonos absolutos.

Por lo que al nivel fonológico profundo se refiere y, sobre todo, en lo que atañe a los lazos que existen entre el componente fonológico y los componentes superiores, no se ha llevado a cabo sino algún estudio piloto que no permite considerar todavía que dichos aspectos forman parte del estado actual del modelo. Sin embargo, como ya hemos indicado, los autores prevén esta posibilidad y empiezan a indagar en este ámbito (véase Portes, Rami, Auran y Di Cristo, 2002).

5.5. Conclusiones

Así, el modelo de Aix-en-Provence que, como hemos visto, pretende llegar a ser en un futuro un modelo completo de la prosodia, constituye una teoría fonológica de la entonación que presenta en su estado actual un interés evidente para los entonólogos, a pesar de las dificultades que existen todavía y que acabamos de mencionar.

En efecto, dicho modelo puede dar cuenta de la organización melódica de los enunciados de lenguas muy distintas y, por consiguiente, permite y facilita el

³⁵ [Hirst, Di Cristo y Espesser \(2000\) presenta la situación del modelo más actualizada y completa e incluye propuestas de futuras implementaciones en el modelo.](#)

³⁶ [Actualmente en el Laboratoire Parole et Langage de la Universidad de Aix-en-Provence se están llevando a cabo estudios que permitan en un futuro próximo incluir en el programa un dispositivo para segmentar automáticamente los enunciados en unidades entonativas.](#)

VERSIÓN BORRADOR PROVISIONAL

comparatismo lingüístico. Esto se ha puesto de manifiesto en Hirst y Di Cristo (1998), publicación en la que se ha aplicado la misma metodología para describir las características prosódicas de veinte lenguas muy diversas (español, inglés, finlandés, árabe, chino mandarín, tailandés, etc.), y que ha desembocado en una clasificación de las lenguas en función de cómo se estructuran las distintas unidades prosódicas fundamentada en criterios unificados.

Por otra parte, una de las peculiaridades innovadoras del modelo de Aix-en-Provence es su enfoque: contrariamente a la mayoría de teorías existentes, permite no sólo generar patrones acústicos de la melodía a partir de una entrada simbólica, sino también llevar a cabo el proceso inverso, que consiste en obtener una codificación de la curva de F0 a partir de la señal acústica, por lo que se facilita el estudio de las relaciones que existen entre los niveles acústicos y fonológicos de la entonación.

Todo ello se apoya además en unos algoritmos altamente automatizados que, aunque deben en algún caso ser completados, constituyen una de las ventajas del modelo, por cuanto reducen el grado en que el análisis depende del investigador, permiten llegar a resultados más fácilmente comparables y generalizables y dar cuenta, como hemos visto, de la relación que existe entre la realidad acústica de los enunciados y la estructuración fonológica de su entonación.

Así, como se habrá podido observar, el modelo de Aix-en-Provence, si bien comparte con otros ciertas bases teóricas y aunque todavía no pueda considerarse del todo completo, constituye una teoría no sólo coherente en el estado actual de desarrollo, sino también original por alguno de sus enfoques.

5.6. Referencias bibliográficas

Alcoba, Santiago y Julio Murillo (1998). Intonation in Spanish. En Hirst, Daniel J. y Albert Di Cristo (eds.) (1998). *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press. 152-166.

Alcoba, Santiago, Martine Le Besnerais y Julio Murillo (1993). Unité tonale et structure prosodique de l'espagnol. *Revue de Phonétique Appliquée* **105**. 261-285.

Astesano, Corine, Robert Espesser, Daniel J. Hirst y Joaquim Llisterri (1997). Stylisation automatique de la fréquence fondamentale: une évaluation multilingue. En *Actes du 4^{ème} Congrès Français d'Acoustique*. Marseille: Société Française d'Acoustique. 441-443.

Bolinger, Dwight (1964). Intonation as a universal. En Lunt, Horace G. (ed.) *Proceedings of the 9th International Congress of Linguists*. The Hague: Mouton. 833-844.

Bruce, Gösta (1977). Swedish word accents in sentence perspective. *Travaux de l'Institut de Linguistique de Lund* **12**. 1-155.

Bruce, Gösta (1989). Report from the IPA working group on suprasegmental categories. *Working Papers* **35**. 25-40.

Campbell, W. Nick (1992). *Multi-level Timing in Speech*. Tesis doctoral. Sussex University.

VERSIÓN BORRADOR PROVISIONAL

- Campione, Estelle, Daniel J. Hirst y Jean Véronis (2000). Stylisation and symbolic coding of F0 : comparison of five models. En Botinis, Antonis (ed.) (2000). *Intonation: Models and Theories*. Dordrecht: Kluwer Academic Publishers. 185-208.
- Campione, Estelle, Emmanuel Flachaire, Daniel J. Hirst y Jean Véronis (1997). Stylisation and symbolic coding of F0, a quantitative approach. En Botinis, Antonis, Georgios Kouroupetroglou y George Carayiannis (eds.) (1997). *Intonation : Theory, Models and Applications*. Athens: ESCA/University of Athens. 71-74.
- Cohen, Antonie y John 'T Hart (1965). Perceptual analysis of intonation patterns. En Commins, Daniel E. (ed.). *Proceedings of the 5th International Congress on Acoustics*. Liège. Paper A16.
- Collier, René (1985). The setting and resetting of the baseline. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* (Tokyo) **19**. 111-132.
- Cooper, William E. y John M. Sorensen (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America* **62**, **3**. 683-692.
- Cooper, William E. y John M. Sorensen (1981). *Fundamental frequency in sentence production*. New York/Heidelberg/Berlin: Springer.
- de Pijper, Jan Roelof (1979). Close-copy stylisation of British English intonation contour. *IPO Annual Progress Report* **14**. 66-71.
- Di Cristo, Albert (1982). *Prolégomènes à l'étude de l'intonation. Micromélogie*. Paris: CNRS.
- Di Cristo, Albert (1985). *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence: Publications de l'Université d'Aix-en-Provence.
- Di Cristo, Albert y Daniel J. Hirst (1997). L'accentuation non-emphatique en français: stratégies et paramètres. En Perrot, Jean (ed.). *Polyphonie pour Ivan Fónagy*. Paris: L'Harmattan. 71-101.
- Drommel, Raimund H. (1980). Towards a subcategorization of speech pauses. Temporal variables in speech. En Dechert, Hans W. (ed.). *Studies in honour of Frieda Goldman-Eisler*. The Hague : Mouton. 227-238.
- Espesser, Robert (1982). Un système de détection du voisement et de F0. *Travaux de l'Institut de Phonétique d'Aix* **8**. 241-261.
- Espesser, Robert (1996). *User's Guide to Prosody Tools*. http://www.lpl.univ-aix.fr/projects/mes_signaix.htm/mes_signaix.tutorial.html.
- Estruch Axmacher, Mònica (1999): *Anàlisi melòdica i codificació simbòlica d'un corpus de paràgrafs en català*. Trabajo de investigación de Tercer Ciclo. Universitat Autònoma de Barcelona.
- Fujisaki, Hiroya (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. En MacNeilage, Peter F. (ed.). *The production of speech*. New York: Springer Verlag. 39-55.
- Garrido, Juan María (1993). Analysis of Global Pitch Contour Domains at Paragraph Level in Spanish Reading Text. *Working Papers* **41**. 104-107.
- Garrido, Juan María (1996). *Modelling Spanish Intonation for Text-to-Speech Applications*. Tesis doctoral. Universitat Autònoma de Barcelona. URL: <http://liceu.uab.es/~juanma/tesis.html>.

VERSIÓN BORRADOR PROVISIONAL

- Grønnum Thorsen, Nina (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America* **77**, **3**. 1205-1216.
- Grønnum Thorsen, Nina (1986). Sentence Intonation in textual context - supplementary data. *Journal of the Acoustical Society of America* **80**, **4**. 1041-1047.
- Hermes, Dik J. y J.C. van Gestel (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America* **90**. 97-102.
- Hirst, Daniel J. (1999). The symbolic coding of duration and timing: an extension to the INTSINT system. *Proceedings of Eurospeech '99*. Budapest, Hungría. 1639-1642.
- Hirst, Daniel J. y Albert Di Cristo (1984). French intonation: a parametric approach. *Die Neueren Sprachen* **83**, **5**.
- Hirst, Daniel J. y Albert Di Cristo (eds.) (1998). *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Hirst, Daniel J. y Robert Espesser (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* **15**. 71-85.
- Hirst, Daniel J., Albert Di Cristo, Martine Le Besnerais, Zora Najim, Pascale Nicolas y Pascal Roméas (1993). Multi-lingual modelling of intonation patterns. *Working Papers* **41**. 204-207.
- Hirst, Daniel J., Albert Di Cristo y Robert Espesser (2000): Levels of representation and levels of analysis for the description of intonation systems. En Horne, Merle (ed.). *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Press. 51-87.
- Hirst, Daniel J., Pascale Nicolas y Robert Espesser (1991). Coding the F0 of a continuous text in french: an experimental approach. *Proceedings of the 12th International Congress of Phonetic Sciences*. Aix-en-Provence. 234-237.
- Ladd, D. Robert (1984). Declination: a review and some hypotheses. *Phonology Yearbook* **1**. 53-74.
- Le Besnerais, Martine (1995). *Contribution à l'étude des paramètres rythmiques de la parole. Analyse contrastive de réalisations phoniques en espagnol et en français*. Tesis doctoral. Universitat Autònoma de Barcelona.
- Lehiste, Ilse (1975). The phonetic structure of paragraphs. En Cohen, Antoine y Sieb G. Nooteboom (eds.). *Structure and Processes in Speech Perception*. Heidelberg: Springer Verlag. 195-203.
- Lieberman, Philip (1967). *Intonation, perception and language*. Cambridge: MIT Press.
- Llisterri, Joaquim (ed.) (1996). *Prosody Tools Efficiency and Failures*. WP 4 Corpus. T4.6 Speech Markup and Validation. Deliverable 4.5.2. Final version. 15 October 1996. LRE Project 62-050 MULTEXT.
- McAllister, Robert (1971). Predicting physical aspects of English stress. *Stockholm Transmission Laboratory Quarterly Progress Status Report* **1**. 20-29.
- Mora, Elsa (1996). *Caractérisation prosodique de la variation dialectale de l'espagnol parlé au Vénézuéla*. Tesis Doctoral. Université d'Aix-en-Provence.
- Nicolas, Pascale (1995). *La structuration prosodique du texte lu en français*. Tesis doctoral. Université de Provence.

VERSIÓN BORRADOR PROVISIONAL

- Nicolas, Pascale y Daniel J. Hirst (1995). Symbolic coding of higher level characteristics of fundamental frequency curves. *Proceedings of the 4th European Conference on Speech Communication and Technology*. Madrid. 989-992.
- O'Shaughnessy, Douglas (1976). *Modelling fundamental frequency and its relationship to syntax, semantics, and phonetics*. Tesis doctoral. MIT, Cambridge (Mass.).
- Pierrehumbert, Janet (2000). Tonal elements and their alignment. En Horne, Merle (ed.). *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Press.
- Pierrehumbert, Janet y Mary Beckman (1988). *Japanese tone structure*. Cambridge: MIT Press.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Portes, Cristel, Emmanuelle Rami, Cyril Auran y Albert Di Cristo (2002). Prosody and Discourse: a Multi-linear Analysis. *Proceedings Speech Prosody 2002*. Aix-en-Provence. 579-582.
- Riera, Montserrat (2001). *Anàlisi acústica dels moviments tonals del grup accentual en català*. Trabajo de investigación de Tercer Ciclo. Universitat Autònoma de Barcelona.
- Rossi, Mario (1999). *L'Intonation, le système du français: description et modélisation*. Paris: Ophrys.
- Rossi, Mario, Albert Di Cristo, Daniel Hirst, Philippe Martin, y Yukihiro Nishinuma (1981). *L'intonation. De l'acoustique à la sémantique*. Paris: Klincksieck.
- Sorensen, John M. y William E. Cooper (1980). Syntactic coding of fundamental frequency in speech production. En Cole, R.A. (ed.). *The perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum Associates. 399-440.
- Terken, Jacques y Dik J. Hermes (2000). The perception of prominence. En Horne, Merle (ed.). *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Press.
- 'T Hart, John (1991). F0 stylisation in speech: straight lines versus parabolas. *Journal of the Acoustical Society of America* **90**, **6**. 3368-3370.
- Thorsen, Nina (1980). Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual Report of the Institute of Phonetics University of Copenhagen* **14**. 1-29.