Modelo cuantitativo de entonación del español

David Escudero Mancebo y **Valentín Cardeñoso Payo**
{descuder,valen}@infor.uva.es
Universidad de Valladolid
Valladolid, España

Resumen En este trabajo se presenta un modelo de entonación basado en funciones de Bézier y patrones estadísticos. Se presentan brevemente diversos modelos de entonación, describiendo la funcionalidad y objetivos de diseño comunes a todos ellos, y las particularidades e inconvenientes que presentan. Los primeros resultados del uso de nuestro modelo indican que aporta ventajas tanto desde el punto de vista del análisis automático, como desde el punto de vista de la mejora de la naturalidad. Por último se detallan las características de un corpus necesario para aplicar el modelo de entonación en una tarea de lectura de noticias.

1 *Introducción*

Uno de los desafíos de las tecnologías del habla de los últimos años es el cambio de las características del locutor en conversores texto voz (CTV) de forma económica. La entonación es uno de los aspectos más relevantes a la hora de imitar el habla de una persona o el de determinadas situaciones y contextos. Ya existen métodos que permiten abordar estos problemas con buenos resultados. Sin embargo, presentan limitaciones importantes que deben ser superadas.

El principal problema de algunos de ellos, es que requieren de gran cantidad de muestras que deben ser procesadas y etiquetadas cuidadosamente, (en ocasiones a nivel segmental) para poder obtener resultados de calidad (ver p.e. [22]).

Otros se apoyan esencialmente en la sílaba acentuada como principal modificador de los perfiles de entonación. No está claro que sea éste el principal evento prosódico del español, donde se han estudiado otras unidades de entonación como el grupo acentual y el grupo de entonación (ver p.e. [11]).

Se constata la necesidad de una codificación efectiva de los patrones de entonación. Es necesario encontrar una representación adecuada de los perfiles de entonación que permita recoger los movimientos importantes y que permita afrontar métodos de reconocimiento de patrones y de generación automática de los mismos. La estilización de segmentos rectos por ejemplo no parece la forma más eficaz de representar contornos superiores a la sílaba [13].

Nosotros proponemos parametrizar los contornos de $F0$ empleando el grupo acentual como unidad básica de referencia. A cada grupo acentual se le asigna un patrón de entonación que se parametriza con una función de Bézier. De este modo se obtienen ventajas tanto desde el punto de vista de la inferencia automática de los patrones de entonación característicos como de la generación automática de los mismos.

Primero se describen otros modelos de entonación descritos en la bibliografía apuntando las principales limitaciones de éstos. Después se presentan las características principales de la nueva propuesta, describiendo su uso en reconocimiento de patrones de entonación y en síntesis de los mismos. Por último se detalla la forma y contenidos de un corpus que se está elaborando y que servirá para mostrar con mayor solidez la validez del método.

2 *Modelos de Entonación*

El estado del arte actual sobre modelos de entonación presenta disparidad de modelos y métodos que afrontan el problema con mayor o menor éxito. Para poder comparar dichos modelos entre sí y nuestra propuesta con los modelos existentes, es necesario definir un marco o esquema que caracterice el funcionamiento de dichos modelos. También es importante establecer cuáles son los objetivos de diseño de un buen modelo de entonación. Ambos aspectos serán tratados en la siguiente sección. Apoyándonos en el esquema definido, y con los objetivos de diseño presentes, se comentan algunos de los modelos de entonación existentes remarcando los aspectos de dichos modelos que pueden ser mejorables, lo que dará pie a presentar y defender el modelo de entonación que presentamos.

2.1 Características Generales

El esquema de la figura 1 representa una abstracción del funcionamiento de un *Módulo Generador de Entonación* (MGE) dentro de un CTV. A partir de un texto, y teniendo en cuenta las características del locutor que lo va a pronunciar, se obtiene un perfil de entonación.

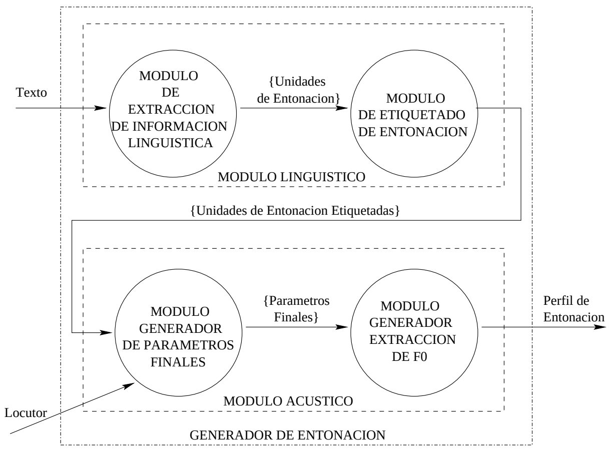

Figura 1: *Esquema funcional de un Módulo Generador de Entonación en un Sistema Traductor Texto Voz.*

En el módulo MGE se distinguen dos componentes menores: un *Módulo Lingüístico* (ML) y un *Módulo Acústico* (MA). ML se encarga de interpretar el texto y extraer de él *Unidades de Entonación Etiquetadas* (UEE), que son componentes lingüísticas representativas desde el punto de vista de la entonación. Estas UEE tienen asociada información sobre su significado prosódico. Los factores dependientes del locutor modulan el funcionamiento del MA pero no del ML.

En el ML se distinguen dos tareas distintas: la extracción de información lingüística y el etiquetado de entonación. La primera tarea se realiza en un módulo de extracción de información lingüística (MEIL). Este módulo puede ser tan complejo como un módulo de procesamiento de lenguaje natural que analice la estructura sintáctica o tan simple como un silabicador-acentuador. En todo caso, a la salida de este módulo se obtendrán un conjunto de *Unidades de Entonación* (UE). Las UE, que pueden ser más o menos complejas, tienen asociados una serie de atributos con información sobre sus características. Estos atributos pueden hacer referencia tanto a las unidades aisladas como a la relación de las UL con UL vecinas. Si empleamos la sílaba como unidad de entonación, estas unidades tienen atributos referentes a su forma (e.g. número de fonemas, abierta ...) y atributos que hacen referencia a su relación con otras sílabas (p.e. sinalefas, fin de palabra ...). Las UE pueden organizarse jerárquicamente para cubrir unidades de estudio mayores con significado prosódico (p.e. grupos acentuales, grupos de entonación...) La segunda tarea se realiza en un módulo de etiquetado de entonación (MEE), cuya función es enriquecer las unidades de entonación caracterizándolas desde el punto de vista de la función que realizan en la entonación. En este módulo se resuelve la relación entre la lingüística y la entonación.

Este módulo acústico se descompone a su vez en otros dos menores. De un lado tenemos un *Módulo Generador de Parámetros Finales* (MGPF) y después un *Módulo Generador de F0* (MGF). El MGPF, genera *Parámetros Finales* (PF) a partir de las UEE. Para generar los PF a partir de las UE, el MGPF tiene en cuenta un conocimiento sobre el locutor.

Por último, el *Módulo Generador de F0*, traduce los parámetros finales en una nube de puntos $(t_i, F0_i)$, que es el perfil de entonación sintético. La función que traduce los parámetros finales en la curva de F0 es explícita y no depende ya de ninguna representación sobre el conocimiento de relaciones lingüísticas o del locutor.

2.2 Objetivos de Diseño

Tal y como se detalla en [24], los objetivos principales de diseño de todo modelo de entonación son:

1. Capacidad de sintetizar perfiles de entonación a partir de una representación textual lingüística.

2. Las representaciones internas del modelo deberían poder generarse automáticamente a partir de los perfiles de entonación.

Aparte de estos requisitos principales, la representación de la entonación debe ser lo más sencilla posible, reduciendo al máximo el número de parámetros empleados. También es importante que las representaciones empleadas puedan cubrir el mayor número posible de movimientos representativos en los perfiles de entonación. Los parámetros empleados deben ser interpretables, de manera que se les pueda asociar un significado lingüístico desde el punto de vista de la entonación.

A estas cualidades apuntadas en [24], es necesario añadir por consideraciones prácticas, que el número de parámetros asociados a las distintas unidades de entonación sea fijo para facilitar así la comparación entre ellas. Además, el sistema de representación deberá facilitar el uso de medidas de distancia entre unidades que reflejen la diferencia perceptual entre patrones de entonación. Por último cuanto menor sea el coste de ajuste a nuevos locutores y/o situaciones mejor.

2.3 Propuestas Existentes

Los modelos de entonación que se usan en traducción texto voz, detallan, sobre el esquema descrito anteriormente, los atributos de las unidades intermedias así como el modo en que los distintos módulos resuelven su funcionalidad. A continuación se describe brevemente cada uno de ellos, haciendo hincapié en sus principales limitaciones.

En los **Modelos Basados en Tonos** las etiquetas de entonación asignadas a cada unidad de entonación son tonos. Se impone un modelo que utiliza solamente dos categorías de tonos **High H** y **Low L** [18]. *ToBI* es un sistema de transcripción prosódica aceptado internacionalmente [20], que establece una serie de reglas para asignar tonos a las unidades de entonación. Se han descrito métodos que asignan automáticamente tonos a las distintas sílabas [27] que aparecen en una locución, y también para generar perfiles de entonación a partir de texto enriquecido con etiquetas ToBI [1] [3] [2]. Sosa[21] realiza una descripción de la entonación del español empleando estos métodos, sin embargo, no considera la correpondencia que lleva de una representación tonal al perfil de entonación. Los métodos referidos anteriormente se apoyan en abundante material etiquetado prosódicamente con marcas ToBI. No existe un corpus similar para el español. Con respecto a las inflexiones del perfil de F0 en unidades pretonemáticas, se hacen recaer principalmente en los acentos, algo que no es tan claro en español.

El **Modelo IPO** describe los perfiles de entonación como *movimientos del pitch*. Sólo se representan los cambios que son perceptualmente relevantes en el perfil de entonación. Los perfiles se aproximan con segmentos rectos que se unen en los movimientos del pitch. La entonación de una lengua viene descrita con una gramática de movimientos [13]. Garrido [11] desarrolla un estudio de la entonación del español siguiendo este método. Encuentra una serie de patrones característicos a nivel de grupos acentuales y estudia los efectos de la declinación en los grupos de entonación. El principal problema de su enfoque es que no está clara la correspondencia entre el texto y los movimientos del pitch. También es un problema el hecho de que los patrones de entonación no tienen todos el mismo número de parámetros. López [16] evita estos problemas aplicando la estilización a nivel de sílaba. El perfil de $F0$ en el núcleo vocálico de cada sílaba se parametriza con tres puntos unidos por segmentos. Esto posibilita la creación de una base de datos prosódica donde las sílabas se categorizan en función de diversos valores prosódicos. Al tener que considerar gran número de tipos de unidades de entonación y la posición de la sílaba en su interior, el número de tipos de sílaba se multiplica y con él el tamaño de la base prosódica. El problema se soluciona empleando procedimientos de clustering para reducir el tamaño de la base prosódica [15]. Sin embargo, el trabajo de asignación de perfiles estilizados a las sílabas es costoso, ya que implica segmentar un corpus, aproximar los perfiles de F0 y comprobar la corrección de dichas aproximaciones retocando los perfiles si es necesario.

En los **Modelos de Superposición** (los más representativo *Fujisaki* y otros [10] [25]), una serie de comandos secuenciados en el tiempo producen por composición el perfil de entonación final. Aunque sean los modelos que más se aproximan al modelo de producción fisiológico de la entonación, resulta costoso encontrar automáticamente los valores de los parámetros del modelo a partir de un corpus.

El **Modelo Tilt**, define cuatro unidades de entonación básicas: acentos tonales, acentos frontera, conexiones y silencios. Los dos primeros se codifican empleando aproximaciones parabólicas y reduciendo los parámetros de éstas a parámetros *Tilt* más intuitivos y representativos de la evolución del pitch [24]. Para este modelo se han definido métodos de análisis automático y de síntesis que ofrecen buenos resultados en entonación del inglés [23] [5]. Este modelo asigna patrones de entonación a los distintos núcleos silábicos de las sílabas acentuadas o frontera. No está claro el papel principal de estas sílabas en los patrones de entonación del español.

El **Modelo INTSINT** presenta un sistema de etiquetado prosódico cuyo objetivo es realizar una descripción de la configuración del perfil de entonación, sin pretender interpretarlo. Se aproxima el perfil de F0 con un spline cuadrático y sobre él se establecen marcas prosódicas a 2/3 de la duración de las palabras. Las etiquetas marcan si hay máximos, mínimos, subidas y bajadas. A la hora de sintetizar estas curvas a partir del texto se describen métodos estadísticos [14] [26] La aproximación cuadrática, aún siendo mejor que la estilización IPO, pierde gran cantidad de detalles del perfil de F0. No está claro en qué posición de las unidades de entonación se asignan las etiquetas para el caso del español. Además, el MGPF no es una función invertible lo que hace dudar de la precisión de los contornos sintéticos.

3 *Modelo Alternativo basado en Funciones de Bézier*

La hipótesis inicial de nuestro modelo es que cada tipo de grupo acentual tiene asociado una clase de patrón de entonación. Patrones de entonación asociados a un mismo tipo de grupo acentual tienen formas similares. Los patrones de entonación marcan la evolución del perfil de F0 a lo largo del grupo acentual, y serán aproximados por funciones de Bézier [8].

El uso funciones de Bézier trae consigo una serie de ventajas de cara a la codificación efectiva de los perfiles de entonación:

- Cada función se parametriza con un número fijo de parámetros.
- La calidad del ajuste se puede ajustar en función del grado del polinomio empleado: a mayor grado, mayor calidad.
- Existen procedimientos de ajuste automático de una nube de puntos a funciones de Bézier [19].

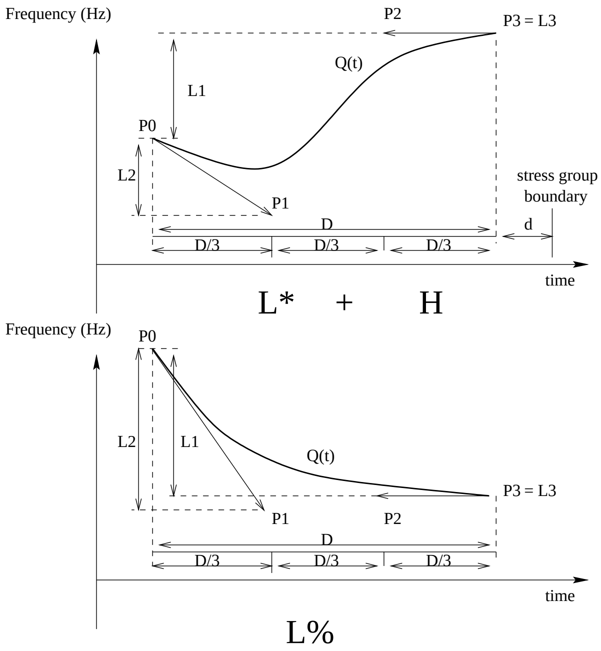

Figura 2: *Conjunto de parámetros para dos patrones de entonación diferentes.*

- La forma de las funciones de ajuste puede estar restringida a determinadas configuraciones.
- Desde un punto de vista gráfico, los parámetros de la curva son muy intuitivos y pueden combinarse entre ellos buscando nuevos parámetros más representativos desde el punto de vista lingüístico, sin perder poder descriptivo.

La figura 2 muestra funciones de Bézier para dos tipos diferentes de patrones de entonación. La función de Bézier de la parte superior de la figura, puede representar un grupo acentual con un acento tonal de la forma L*+H. En la parte inferior, se representa un patrón de entonación de un tono frontera del tipo L%. Ambas curvas tienen grado tres con puntos de control $P0$, $P1$, $P2$ y $P3$. Los puntos de control y la duración, D, determinan la forma de la función. Se permite que los patrones de entonación estén desplazados con respecto al final del grupo acentual. Los parámetros $L1$, $L2$ y $L3$ (ver figura) se calculan a partir de los puntos de control, y constituyen un juego de parámetros con un mayor significado lingüístico.

Para un mismo juego de parámetros, podemos obtener patrones diferentes variando los valores de dichos parámetros. Si fuera necesario describir un máximo local en el patrón de entonación, basta añadir un grado más al

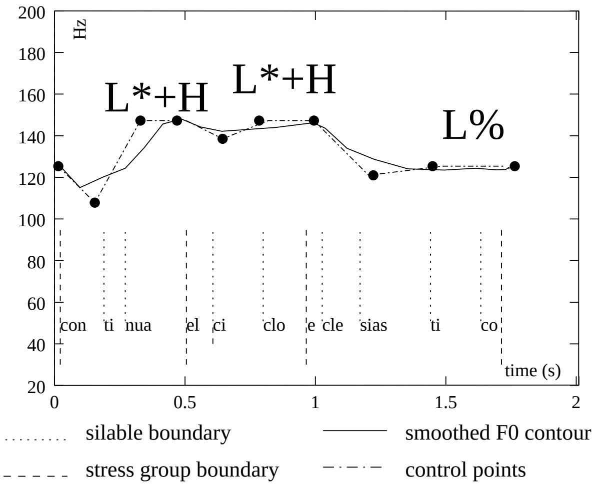

Figura 3: *Perfil de F0 suavizado y los puntos de control de lasj funciones de Bézier que lo aproximan.*

polinomio de la función de Bézier.

En los siguientes apartados, se muestra cómo se puede aplicar de esta técnica en análisis y síntesis automática de la entonación.

3.1 Análisis Automático

El modelo permite adaptarse a nuevos locutores y/o situaciones. Para ello, se identifican los patrones de entonación en un corpus y son agrupados en clases en función de su forma. Una vez generada la base de datos de clases de patrones de entonación, se infieren los modelos que representan cada clase. El trabajo de análisis se desarrolla pues en dos etapas: la identificación de los patrones de entonación y la posterior clasificación. Estos métodos han sido ya descritos en [6], por lo que aquí sólo haremos un apunte sobre su funcionamiento.

Para identificar patrones empleamos un método de ajuste de una nube de puntos (en este caso el perfil de F0) con un conjunto de funciones de Bézier (en este caso los patrones de entonación). El criterio que dirige el ajuste es el de minimizar la distancia de las funciones de Bézier resultantes con la nube de puntos. Un método de programación dinámica establece las fronteras entre funciones de Bézier consecutivas. Se controla la forma de las funciones de Bézier resultantes para que se aproximen a las formas generales conocidas de los patrones de entonación del español (ver figura 3).

Una vez que se obtiene la secuencia de patrones de entonación, éstos son parametrizados y clasificados de acuerdo a técnicas de clustering clásicas [4]. Se aplica una función de máxima similitud como criterio de pertenencia.

En [7] se detalla la aplicación concreta del método a la parametrización de un corpus sencillo de 24 frases enunciativas cortas pronunciadas por dos locutores diferentes. Los patrones de entonación elegidos son de la forma $L * +H$ y $H*$ para los pretonemáticos y $L\%$ para los tonemáticos (ver [21]). Son localizados correctamente un porcentaje superior al 90% de los grupos acentuales en el caso de grupos acentuales iniciales y finales.

Elegimos utilizar tres clases de patrones de entonación: una para los grupos acentuales iniciales, otra para los intermedios y otra para los finales. Cada clase se representa por el vector de medias $\bar{\mu}$ y la matriz de covarianza $\sum$ de los valores de los parámetros $L1$, $L2$ y $L3$ de los patrones de entonación de dicha clase. Aplicando un test de normalidad a los valores de los parámetros en las distintas clases, sólo se producen fallos en los valores de algún parámetro correspondiente a los patrones de entonación centrales. Esto es debido a que en estas zonas es donde se produjeron el mayor número de fallos al identificar los patrones de entonación.

Se aplicó una experiencia de clasificación donde dado un nuevo patrón de entonación de un determinado tipo, se determinada a qué clase asignarlo. Los resultados fueron próximos al 100% cuando se comparaban los grupos iniciales y los finales.

3.2 Síntesis Automática

Para generar perfiles de entonación sintéticos, primero se asigna una clase de patrón de entonación a cada grupo acentual, después se genera el patrón correspondiente y por último se concatenan los patrones de grupos acentuales consecutivos.

Cada clase de patrón de entonación tiene asociada una distribución multivariable $N(\bar{\mu}, \sum)$. Los valores de $\bar{\mu}$ pueden ser utilizados para los parámetros $L1$, $L2$ y $L3$. Otra opción consiste en utilizar un método de generación de números aleatorios que se ajusten a dicha distribución [12]. El valor del parámetro D, asociado a cada grupo acentual se establece en un módulo de duración independiente basado en [17].

El método ha sido probado utilizando una herramienta PSOLA. Ambas opciones fueron satisfactorias desde un punto de vista per-

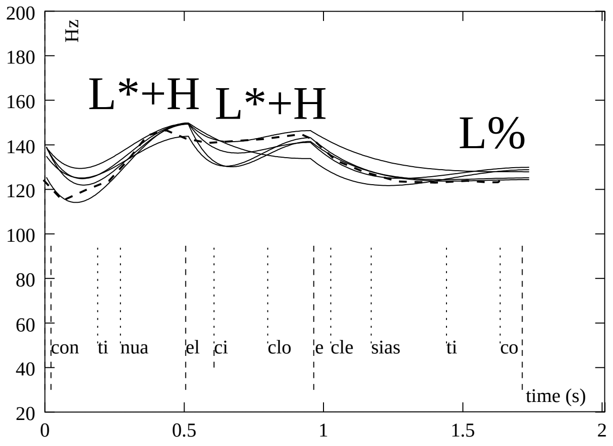

Figura 4: *Diversos contornos de F0 generados para la misma frase. La línea discontinua muestra el perfil F0 original.*

ceptual. Se ha observado que la opción de generar valores de parámetros aleatorios evita la monotonía que se percibe al sintetizar un número elevado de frases. Se ha observado que cuando una misma frase es sintetizada empleando contornos de F0 diversos, se perciben las diferencias aunque todas las locuciones se perciben como naturales (ver figura 4).

El modelo también ha sido incluido en un sistema CTV propio [9]. Se comparan las producciones obtenidas con las obtenidas con una implementación propia del modelo de entonación descrito en [16]. Aunque los resultados son similares, se observa que se produce voz menos monótona cuando el número de frases se incrementa.

4 *Elaboración de un Corpus*

Para avalar el funcionamiento del método y entrenarlo y probarlo de manera que abarque un mayor número de patrones y pueda ser empleado en un amplio rango de situaciones, se requiere de un corpus más completo que el empleado en la primera experiencia.

La tarea elegida es la de lectura de noticias de prensa. El uso de sistemas CTV en esta tarea presenta el problema de que las locuciones acaban resultando monótonas para el oyente porque los patrones de entonación se repiten con frecuencia. Nuestro sistema no aplica patrones de entonación idénticos sino que éstos varían de acuerdo a una distribución estadística. Por esto, es una tarea especialmente adecuada para probar el modelo.

Con respecto al locutor, se elige un locutor varón y de edad joven ya que éstas son las características del locutor que se empleó para crear el módulo de síntesis del sistema CTV que empleamos. La tarea condiciona la elección de un locutor profesional.

Con respecto a los contenidos prosódico-sintácticos del corpus, asumimos de antemano que no vamos a abarcar todos los posibles tipos de frases, de sintagmas, de grupos de entonación que se pueden dar en los textos periodísticos. Por cuestiones prácticas, optamos por distinguir a priori un número limitado de tipos de grupos acentuales lo suficientemente representativos. Con este subconjunto será posible justificar las posibilidades del modelo de entonación presentado. Además, si se eligen los tipos de grupos acentuales adecuados, podrán usarse los modelos entrenados para generar contornos de entonación naturales para un gran número de casos. Así, elegimos sólo frases enunciativas con más de un grupo de entonación y que incluyan tanto grupos de entonación ascendentes como descendentes. De los grupos de entonación considerados en [16] sólo se consideran en esta primera aproximación los grupos terminativos (descendente), continuativos e incidentales (ascendentes). Se descartan los grupos contrastivo, vocativo, parentético, yuxtapuesto, apelativo y apelativo pronominal.

Se emplean textos de un boletín electrónico. Se eligen textos lo más "neutrales" posible para evitar interpretaciones emocionales de los mismos. Son textos relativos a convocatorias de cursos, becas, breves, acuerdos de juntas etc... El locutor deberá leer textos con monotonía y evitando incluir énfasis o emoción en sus locuciones. Se elige este estilo neutro porque su aplicación en lectura de otro tipo de noticias o en otro tipo de tareas es plausible.

Se impone la restricción de emplear un mínimo de 50 grupos acentuales de cada tipo empleado: 40 al menos para entrenar el modelo asociado y 10 para hacer pruebas. 40 se considera un número suficiente para aplicar tests de normalidad sobre los datos. Se distinguen grupos acentuales iniciales, intermedios y finales pertenecientes a grupos de entonación ascendentes y descendentes. De este modo será necesario utilizar locuciones que incluyan al menos 300 grupos acentuales. Se eligen 15 textos de la citada publicación, con 70 frases en total y con 278 grupos de entonación, 208 ascendentes y el resto descendentes.

La grabación se realizará en un estudio profesional. Se identificarán las distintas frases, que serán almacenarán y tratadas de forma independiente. Una vez marcados los grupos acentuales y extraídos los contornos de F0, el corpus estará listo para ser tratado automáticamente y obtener los patrones de entonación.

5 Conclusiones

Los primeros resultados muestran la viabilidad del uso de funciones de Bézier para modelar patrones de entonación de los grupos acentuales. La técnica descrita permite un fácil tratamiento cuantitativo de la entonación tanto en tareas de análisis como de síntesis.

Incluso con un corpus de tamaño pequeño se ha observado que se pueden inferir regularidades en los patrones de la entonación. La elección de valores aleatorios en la generación de perfiles de entonación, permite mejorar la naturalidad de las locuciones sintéticas.

Se ha diseñado otro corpus de tamaño mayor, que con un procesamiento mínimo se espera proporcione la posibilidad de avalar el funcionamiento del método y obtener una gama mayor de perfiles de entonación para emplear el modelo en un mayor rango de aplicaciones.

6 Agradecimientos

Este trabajo ha sido realizado en el marco del proyecto VA16/00A, financiado por la Conserjería de Educación y Cultura de la Junta de Castilla y León, cuyo título es *Aplicaciones de las Tecnologias del Lenguaje Hablado al diseño de interfaces multimodales para centros de respuesta telefónica avanzados.*

Referencias

[1] M Anderson, J Pierrehumbert, and M Liberman. Synthesis by rule of english intonation patterns. In *Proceedings of ICASSP 84*, 1984.

[2] J. R. Bellegarda, K. Silverman, and V. Anderson. Statistical Prosodic Modeling: From Corpus Design to Parameter Estimation. *IEEE Transaction on Speech and Audio Processing*, 9(1):52–66, January 2001.

[3] A. W. Black and A.J. Hunt. Generating f0 contours from tobi labels using linear regression. In *Proceedings of ICSLP 96*, 1996.

[4] R. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* John Wiley and Sons, 2000.

[5] KE Dusterhoff. *Synthesizing Fundamental Frequency Using Models Automatically Trained from Data.* PhD thesis, University of Edimburgh, U.K., 2000.

[6] D. Escudero and V. Cardeñoso. Obtencion automatica de modelos de entonacion a partir de un corpus empleando splines y patrones estadisticos: primeros resultados. In *Actas de las I Jornadas en Tecnologias del Habla. Sevilla 2000*, Noviembre 2000.

[7] D. Escudero and V. Cardeñoso. Una experiencia en reconocimiento automático de tipos de unidades melódicas a partir de su perfil de entonación. In *Actas del II Congreso en Fonética Experimental. Sevilla 2001*, Marzo 2001.

[8] G. Farin. *Curves and Surfaces for CAGD.* Cambrige University Press, 4 edition, 1996.

[9] L. Feal, D. Escudero, and V. Cardeñoso. Un modelo arquitectónico para los sistemas texto-voz. In *Actas de las IV Jornadas de Infomatica. Las Palmas de Gran Canaria*, Junio 1998.

[10] H Fujisaki and K Hirose. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of Acoustics Society of Japan*, 5(4):233–242, 1984.

[11] J. M. Garrido. *Modelling Spanish Intonation for Text-to-Speech Applications.* PhD thesis, Facultat de Lletres, Universitat de Barcelona, España, 1996.

[12] J. E. Gentle. *Random Numbers Generation and Monte Carlo Methods (Statistics and Computing).* Springer, 1998.

[13] J. Hart, R. Collier, and A. Cohen. *A perceptual study of intonation. An experimental approach to speech melody.* Cambrige University Press, 1990.

[14] DJ Hirst, N Ide, and J Veronis. Coding fundamental frequency patterns for multilingual synthesis with intsint in the multext project. In *Proceedings of 2nd*

ESCA/IEEE Workshop on Intonation 1994, pages 77–81, Septembre 1994.

[15] E. López, J. M. Rodríguez, L. Hernández, and J. M. Villar. Automatic prosodic modeling for speaker and task adaptation in text-to-speech. In *Proceedings of ICASSP 97*, 1997.

[16] Eduardo López. *Estudio de Técnicas de Procesado Lingüístico y Acústico Para Sistemas de Conversión Texto Voz en Español Basados en Concatenación de Unidades.* PhD thesis, E.T.S.I de Telecomunicaciones, Universidad Politécnica de Madrid, España, 1993.

[17] A. Macarrón, G. Escalada, and M. Á. Rodríguez. Generation of duration rules for a spanish text-to-tpeech synthesizer. In *Proceedings of Eurospeech 91*, 1991.

[18] J. B. Pierrehumbet. *The Phonology and Phonetics of English Intonation.* PhD thesis, MIT, 1980.

[19] Michael Plass and Maureen Stone. Curve-Fitting with Piecewise Parametric Cubics. *Computer Graphics*, pages 229–239, July 1983.

[20] K. Silverman, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. Tobi: A standard for labelling english prosody. In *Proceedings of ICSLP92*, 1992.

[21] J. M. Sosa. *La Entonación del Español.* Cátedra, 1999.

[22] A. Syndal, G. Möeler, K. Dusterhoff, A. Conkie, and A. Black. Three methods of intonation modeling. In *Proceedings of 3rd ESCA Speech Synthesis 1998*, 1998.

[23] P. Taylor. Automatic recognition of intonation from f0 contours using the rise/fall/connection. In *Proceedings of Eurospeech93*, 1993.

[24] P. Taylor. Analysis and Synthesis of Intonation using the Tilt Model. *Journal of Acoustical Society of America*, 107(3):1697–1714, 2000.

[25] J. van Santen and B. Moebius. Modeling pitch accent curves. In *Proceedings of ESCA Workshop on Intonation 1997*, 1997.

[26] J Veronis, P Di Cristo, F Courtois, and C Chaumette. A stochastic model of intonation for text-to-speech synthesis. *Speech Communication*, 26(4):233–244, 1998.

[27] C. W. Wightman and M. Ostendorf. Automatic Labeling of Prosodic Patterns. *IEEE Transaction on Speech and Audio Processing*, pages 469–481, October 1994.