

# Fonética y tecnologías del habla

**Joaquim Llisterri, Lourdes Aguilar, Juan M. Garrido,  
María Jesús Machuca, Rafael Marín, Carme de la  
Mota y Antonio Ríos**

Departament de Filologia Espanyola  
Universitat Autònoma de Barcelona

Edifici B, Universitat Autònoma de Barcelona  
08193 Bellaterra, Barcelona  
{lourdes|juanma|maria|rafa| carme|mestre}@liceu.uab.es;  
joaquim.llisterri@uab.es  
<http://liceu.uab.es/>

## 1. Introducción

La fonética constituye, indudablemente, uno de los ámbitos de la lingüística que se ha beneficiado en mayor medida de la utilización de herramientas computacionales. Tratándose de una disciplina con un importante componente experimental (Llisterri, 1991), es fácil comprender que buena parte de su desarrollo en las últimas décadas haya estado estrechamente ligado a los avances de la informática, tal como en otros momentos lo estuvo a los de la electroacústica o, a finales del siglo pasado, a los de la mecánica. Al igual que en otros campos del saber, disponer de poderosas herramientas de análisis ha favorecido el desarrollo de teorías que, en nuestro caso, tienen como objetivo último explicar la comunicación mediante el habla entre los seres humanos. Dichas teorías avanzan necesariamente gracias al mejor conocimiento de los mecanismos de producción, transmisión y percepción de las señales sonoras lingüísticamente significativas que constituyen el objeto de estudio de la fonética. Sin embargo, la accesibilidad de nuevos sistemas de análisis no garantiza por sí misma el desarrollo de modelos y aunque el presente capítulo presente aspectos fundamentalmente metodológicos y aplicados, no por ello debe olvidarse la existencia de un importante corpus teórico que proporciona una sólida base a las observaciones empíricas (Beckman, 1988; Lindblom, 1990; Stevens, 1989, entre otros).

Por otra parte, la necesidad de transmitir el habla a distancia de un modo eficaz está en la base del surgimiento de las denominadas tecnologías del habla, campo que comprende tanto la síntesis -generación automática del habla a partir de una representación simbólica- como el reconocimiento -conversión del habla en una representación simbólica- integrados en sistemas que permiten establecer una comunicación en forma de diálogo entre la persona y el ordenador (Holmes, 1988; Keller (Ed.), 1994; O'Shaughnessy, 1987). Como es de suponer, el desarrollo de las tecnologías del habla ha sido paralelo al de la informática, especialmente en sus aplicaciones al procesamiento de señales; aun así, el avance de los conocimientos en

fonética ha contribuido también a que este ámbito se encuentre actualmente entre los que ofrecen mayores posibilidades de expansión en el marco de lo que se conoce como las industrias de la lengua (Cole *et al.* (Eds.), 1996; Sager, 1992; Vidal Beneyto (Dir), 1991).

La primera parte de este capítulo presenta sucintamente las principales herramientas informáticas de que dispone la fonética experimental en cada una de sus principales ramas - fonética articulatoria, acústica y perceptiva -, incluyendo algunos ejemplos del método de trabajo y de los resultados obtenidos en nuestro grupo de investigación. La segunda parte aborda la aplicación de los conocimientos derivados mediante estas herramientas a uno de los ámbitos de las tecnologías del habla, la conversión de texto a habla, entendida como un proceso que permite la transformación de un texto escrito en ortografía convencional en su representación sonora; para ello se ofrecen también algunos ejemplos de los trabajos llevados a cabo en el grupo.

El principal objetivo del capítulo es mostrar cómo el uso de herramientas informáticas guiado por una metodología experimental y una teoría que la sostiene permite, por una parte, avanzar en la descripción fonética de las lenguas y, por otra, poner los datos obtenidos al servicio de aplicaciones encaminadas, en primer lugar, a facilitar la interacción entre personas y máquinas y, en última instancia, a establecer un modelo del comportamiento comunicativo humano.

## **2. Herramientas y métodos de análisis en fonética**

Tal como se ha indicado al principio, la fonética es una rama de la lingüística que tiene entre sus objetivos la determinación de las características de los sonidos del habla. Dichas características pueden describirse desde el punto de vista articulatorio, acústico y perceptivo en función de cada uno de los elementos que integran el acto comunicativo -emisor, mensaje y receptor-. Cada configuración articulatoria del emisor se corresponde con unas características acústicas específicas de la señal que el oyente considera a fin de discriminar los indicios que necesita para la percepción o decodificación del mensaje recibido.

El estudio experimental del habla desde las diferentes perspectivas de la fonética requiere la utilización de herramientas de análisis con el fin, por un lado, de establecer una clasificación de los parámetros que caracterizan a los sonidos - o elementos segmentales - de una lengua, y, por otro, de determinar la estructura y el funcionamiento de los patrones entonativos y rítmicos que configuran los denominados elementos suprasegmentales.

El enfoque articulatorio, acústico o perceptivo del estudio fonético condiciona y determina la elección de las técnicas de análisis adecuadas. Algunas de las técnicas empleadas en el estudio de los elementos segmentales son aplicables también al análisis de los suprasegmentales como la entonación, el acento, el ritmo o la velocidad de elocución).

## 2.1. El análisis articulatorio del habla

La aerometría, la electrolaringografía, la electromiografía, la electropalatografía o la radiografía son herramientas que nos permiten abordar el estudio de los procesos implicados en la producción del habla (Abbs-Watkin, 1976; Stone, 1996). Así, para el español, disponemos de la caracterización de los sonidos mediante palatogramas (Navarro Tomás, 1918) y mediante cortes sagitales (Navarro Tomás, 1918; Quilis-Fernández, 1964; Álvarez Henao, 1977; Quilis, 1985).

Dichas técnicas se pueden aplicar no sólo a la descripción de la lengua, sino también a la caracterización y clasificación de los problemas fonéticos en hablantes que padecen algún tipo de alteración del habla o del lenguaje. Por ejemplo, los estudios en palatografía permiten diagnosticar los problemas de articulación, facilitando así la tarea de corrección fonética (Code-Ball, 1984; Ball, 1989).

También el estudio de la melodía, del ritmo y de la calidad de la voz se beneficia de la existencia de instrumentos que analizan la producción, como el laringógrafo, el electroglotógrafo o el estroboscopio .

Por último, cabe destacar que un conocimiento profundo de los mecanismos fisiológicos puede contribuir a la mejora de los modelos de producción (Levelt, 1989) y de los modelos prosódicos (Fujisaki, 1991).

## 2.2. El análisis acústico del habla

El estudio de los segmentos del habla y de los fenómenos de alcance suprasegmental desde una perspectiva acústica implica el análisis y la cuantificación en los dominios de la frecuencia, la amplitud y el tiempo que configuran la onda sonora portadora del habla. Algunas de las técnicas empleadas en esta tarea son el análisis oscilográfico, espectrográfico, espectral y de predicción lineal (LPC) (Javkin, 1996; Wakita, 1996)

En este sentido, Lindblom (1986) se refiere a dos de los problemas abordados desde la fonética acústica: la segmentación y la búsqueda de invariantes acústicos. Para solventar las dificultades de segmentación en unidades fonéticas del continuo acústico que constituye la onda sonora, se recurre generalmente al análisis oscilográfico, dado que es una técnica que ofrece una representación de la señal en los dominios del tiempo y de la amplitud, conocida también como forma de onda. Sin embargo, la delimitación de las fronteras entre los segmentos del habla presenta múltiples problemas. Veamos, por ejemplo, el caso de los sonidos aproximantes. Los sonidos aproximantes, caracterizados articulatoriamente como *made with open approximation of the articulators, and central passage or the air stream* (Abercrombie, 1967: 67), desde el punto de vista acústico son muy similares a las vocales, dada la presencia de periodicidad y de fuerte intensidad. De ahí que la segmentación de una secuencia V-aproximante-V no sea tarea fácil. Como puede observarse en la figura 1, resulta complejo determinar con precisión los límites entre la consonante y las vocales adyacentes. Por otro lado, la corta duración del segmento aproximante acentúa estos problemas. La existencia de tales dificultades, sin embargo, no invalida el trabajo en la caracterización de estos sonidos. Así, disponemos de descripciones temporales de los

sonidos aproximantes del español en Martínez Celdrán (1985) y en Aguilar-Andreu (1991).

El método tradicional de análisis de la duración segmental se basa en la determinación manual de las fronteras por medio de oscilogramas y espectrogramas. Recientemente, sin embargo, se han desarrollado sistemas de segmentación automática que utilizan técnicas de autoaprendizaje, como los Modelos Ocultos de Markov (Boëffard, 1993).

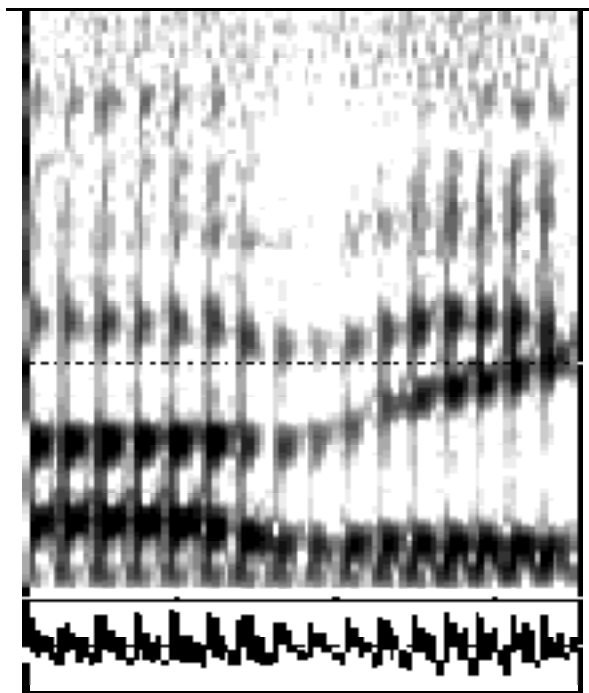


Figura 1. Espectrograma y oscilograma de la secuencia [ˈa]a

En cuanto a la búsqueda de invariantes acústicos, los métodos utilizados - análisis espectrográfico, análisis espectral y análisis por predicción lineal - se centran en la extracción de parámetros frecuenciales, tanto estáticos como en su evolución a lo largo del tiempo.

Mediante el análisis espectrográfico podemos visualizar la onda sonora en los tres dominios del tiempo, la frecuencia y la amplitud. La frecuencia de los formantes, uno de los principales indicios acústicos utilizados como representante de la invariación, se calcula trazando una línea imaginaria en el centro del formante y tomando la frecuencia del punto medio de esta línea (Farmer, 1984; Kent y Read, 1992; Ladefoged, 1996).

En ocasiones, sin embargo, la determinación de la frecuencia de los formantes presenta dificultades. Por ejemplo, en el caso de la consonante nasal palatal, aparece un problema de interpretación de la estructura formántica en lo que concierne a sus formantes segundo y tercero. En la mayoría de espectrogramas se advierte una fusión de los dos formantes, lo que plantea la alternativa de considerar un único formante de banda ancha resultante de la aproximación de dos formantes, o considerar dos formantes de frecuencias muy próximas. En Machuca (1991) se optó por tomar los

datos de dos formantes. La trayectoria de este formante considerando las vocales adyacentes indica que la parte inferior se corresponde con un formante de nasalidad que se crea en las vocales, y la parte superior, con el segundo formante de las vocales (figura 2).

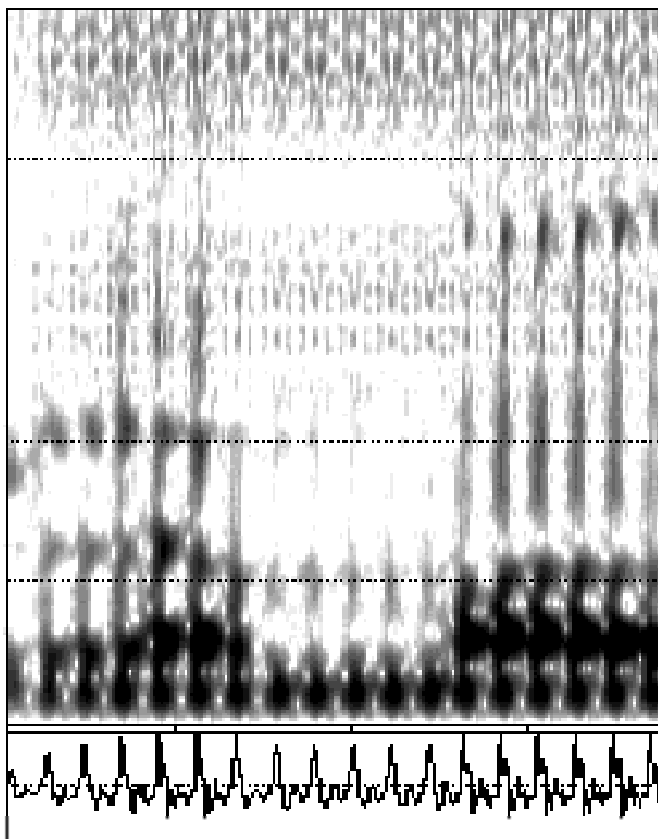


Figura 2. Espectrograma y oscilograma de la secuencia [a'ma]

Al introducir la variable tiempo en la representación de la onda sonora, el espectrograma permite representar la trayectoria de los formantes y por tanto obtener información sobre las transiciones de un sonido a otro (figura 3). Sin embargo, determinar el punto en que se inicia el cambio de frecuencia y el punto en el que se estabiliza de nuevo la trayectoria del formante, es decir, determinar los límites de la transición, es especialmente difícil cuando las muestras de habla proceden de una situación comunicativa informal, caracterizada por una relajación en la pronunciación. Dada esta situación, en el proceso de análisis llevado a cabo en Aguilar (1991), donde el corpus procede de un discurso oral, sin planificación previa, de un informante se optó por considerar la secuencia vocálica como una unidad y tomar los datos de frecuencia de los formantes en cinco puntos equidistantes desde el inicio hasta el fin de la secuencia, de manera que se pudiera automatizar fácilmente el proceso de obtención de datos.

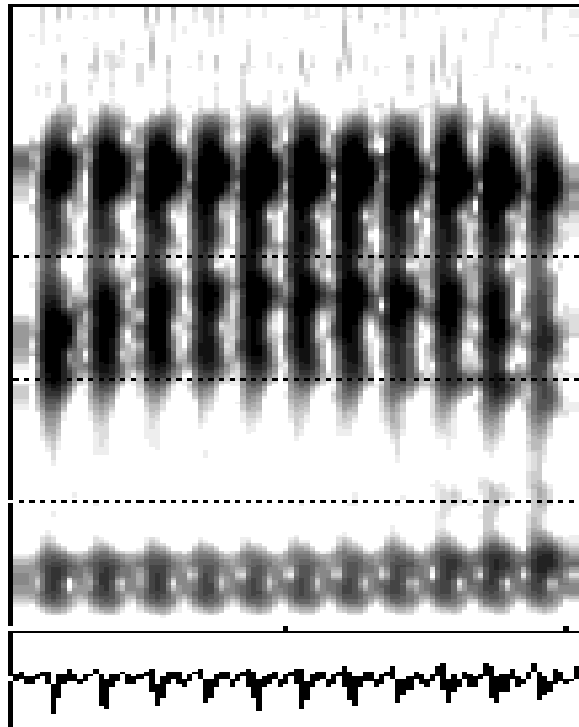


Figura 3. Espectrograma y oscilograma de la secuencia [ja]

Los datos obtenidos a partir del análisis espectrográfico permiten asimismo derivar fórmulas que expliquen la variación fonética: por ejemplo, es posible calcular el índice de reducción de los sistemas vocálicos en diferentes estilos de habla mediante el grado de entropía (Harmegnies-Poch, 1992).

En la determinación de los formantes es posible también utilizar el análisis espectral, mediante el que obtenemos una representación de la onda sonora en los dominios de la frecuencia y la amplitud. El espectro de un sonido es la función resultante de aplicar el algoritmo denominado Transformada Rápida de Fourier (*FFT, Fast Fourier Transform*), que descompone la onda sonora en sus armónicos (Martí, 1988).

Este procedimiento no proporciona información acerca de la evolución frecuencial en el tiempo y, por tanto, no permite analizar las transiciones. Sin embargo, tiene la ventaja de ofrecer un análisis de frecuencias detallado; en otras palabras, es posible observar la distribución de la energía en la escala de frecuencias: así, si el sonido es sonoro, se aprecia la estructura fina de armónicos, y por tanto, la frecuencia fundamental y, si el sonido es sordo, la concentración de la energía en determinadas bandas frecuenciales.

En el espectro, los formantes aparecen en forma de agrupación de armónicos. La frecuencia del formante se toma en el punto medio de la agrupación, mientras que la intensidad corresponde la del armónico más alto. El armónico de frecuencia más baja que se observa en el espectro corresponde a la frecuencia fundamental ( $F_0$ ), que tiene como correlato articulatorio la frecuencia de vibración de las cuerdas vocales. En el modelo de producción del habla de Fant (1960), la parametrización del espectro describe la fuente de excitación en términos de  $F_0$  y el filtro o función de transferencia

del tracto vocal - cuyo correlato articulatorio es la configuración de las cavidades supraglóticas - en términos de las frecuencias de formantes.

Otra técnica que nos permite obtener información sobre la estructura acústica de los sonidos del habla es el análisis por predicción lineal (LPC, *Linear Predictive Coding*), según el cual la onda sonora se representa directamente en términos de parámetros relacionados con la función de transferencia del tracto vocal y las características de la función de la fuente que varían con el tiempo. El análisis LPC se enfoca como un procedimiento de separación entre la estructura fina del espectro y la envolvente espectral - formada por los picos correspondientes a los formantes-, de tal modo que se evitan los problemas inherentes al análisis espectral, como las dificultades en el momento de analizar voz femenina o infantil (Atal-Hanauer, 1971; Atal, 1985).

A partir de los coeficientes LPC se reconstruye, como se ha indicado, la envolvente espectral. Ejemplificamos el uso de esta técnica con la caracterización acústica de los diptongos y hiatos del español presentada en el trabajo de Aguilar (1994). Mediante la sucesión de análisis LPC en una secuencia es posible obtener una representación dinámica de las trayectorias formánticas con una ecuación polinómica de segundo grado  $ax^2 + bx + c$  donde el coeficiente  $a$  equivale al grado y la forma de curvatura. Con el fin de poder comparar segmentos de diferente duración, se lleva a cabo también una normalización temporal. La figura 4 representa el hiato [ía] de forma esquemática siguiendo el procedimiento descrito anteriormente.

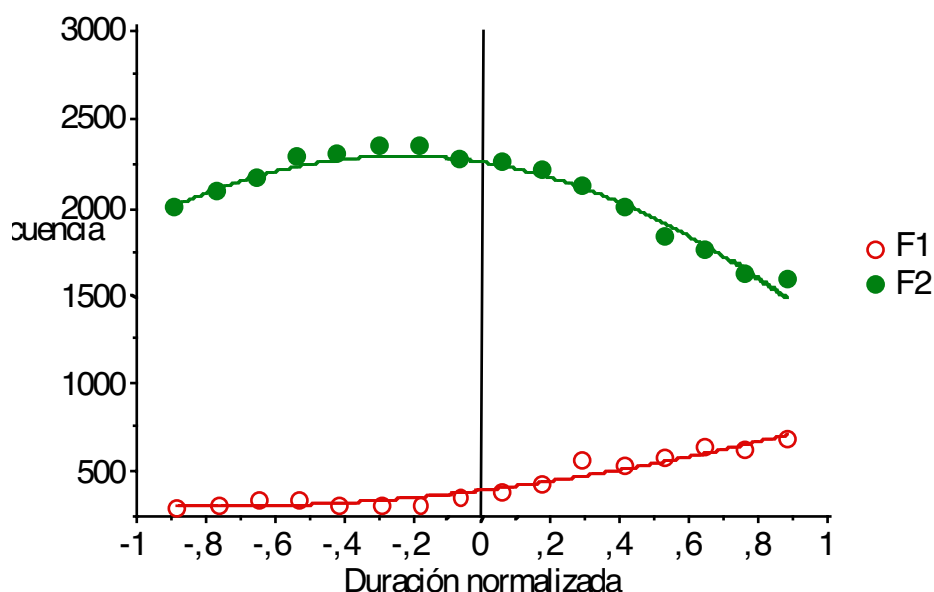


Figura 4. Secuencia [‘ia] extraída de un corpus de frases marco con los puntos de análisis de LPC.

A modo de resumen, cabe mencionar que, en el nivel segmental, la selección de una determinada técnica acústica dependerá del tipo de sonidos que se estén describiendo, de los parámetros que se quieran analizar y de la procedencia de las muestras de habla. Frente al análisis espectral, el análisis espectrográfico presenta las ventajas derivadas de incorporar la información temporal: por ejemplo, posibilita el estudio de las transiciones

y la observación de las trayectorias formánticas. Por otro lado, elimina los efectos de la ventana de análisis, que puede afectar a la caracterización de sonidos cortos, como la vibrante simple del español, o a la descripción acústica de los elementos segmentales en estilos de habla relajados o con una velocidad de elocución elevada donde se dan reducciones fonéticas (de la Mota, 1991; Aguilar *et al.*, 1993; Blecua, 1996).

Por otro lado, la principal diferencia entre el análisis espectral y el análisis por predicción lineal reside en que el segundo método elimina la influencia de la fuente, por lo que no aparece la frecuencia fundamental ni la estructura fina de armónicos, siendo adecuado para generalizar las descripciones de diferentes tipos de voz.

En lo que se refiere al nivel suprasegmental, el objeto principal de análisis es el contorno de la frecuencia fundamental -  $F_0$  o tono - en un dominio determinado. El cálculo del periodo a partir de la forma de onda es el método clásico para el análisis acústico del tono. No obstante, los avances en el procesado digital de señales han traído consigo el desarrollo de algoritmos de estimación de la frecuencia fundamental que emplean técnicas basadas en el análisis de la estructura armónica, en el cálculo LPC o en la detección de picos en el oscilograma (Gold-Rabiner, 1969; Hess, 1983, entre otros). Sin embargo, los contornos frecuenciales obtenidos por estos métodos contienen normalmente errores, que suelen minimizarse por medio de la aplicación de métodos de alisado. Por otro lado, las curvas contienen variaciones muchas veces irrelevantes para el análisis posterior. Así, el estudio comparado de los contornos frecuenciales precisa, por un lado, una estilización o eliminación de las variaciones no relevantes de cada contorno -como las debidas a información micromelódica- y por otro, una normalización o eliminación de las variaciones entre contornos. Recientemente, el desarrollo de sistemas automáticos de estilización ha simplificado enormemente estas tareas (el sistema MOMEL, por ejemplo, descrito en Hirst-Espesser, 1993, entre otros).

Los contornos frecuenciales se han relacionado tradicionalmente con la expresión de la modalidad (Garrido, 1991), el acento (Garrido *et al.*, 1993; Llisterra *et al.*, 1995), la posición dentro del párrafo (Garrido *et al.*, 1993; Garrido, 1996), el contenido informativo (de la Mota, 1995) o la información sintáctica del enunciado (Garrido *et al.*, 1995).

También es posible analizar la amplitud, relacionada habitualmente con la entonación, el acento o el realce fonológico, entre otros, por medio de las denominadas envolventes de energía o de amplitud.

## 2.3. El análisis perceptivo del habla

Los resultados acústicos obtenidos a partir de las técnicas de análisis mencionadas anteriormente pueden validarse perceptivamente, a fin de discernir qué indicios acústicos se utilizan en la percepción del habla. En general, la forma de obtener la opinión de los oyentes consiste en responder a unas determinadas preguntas sobre cada uno de los estímulos que componen una prueba de percepción (Sawusch, 1996).

Los estímulos que se utilizan para la elaboración de las pruebas de percepción pueden obtenerse del habla natural, o bien crearse mediante técnicas de síntesis del habla. La ventaja de utilizar la síntesis en la preparación de estos estímulos es que se pueden

variar independientemente los valores de los parámetros acústicos y, de esta forma, llegar a conclusiones sobre el efecto de los valores de los parámetros en la percepción. Algunos entornos gráficos diseñados para el uso de sintetizadores de habla (Klatt, 1988 en el entorno KPE (*Klatt Parameter Editor*) desarrollado en *University College London*) permiten visualizar de forma conjunta el segmento procedente del habla natural y el segmento sintetizado; así, es posible modificar los parámetros del habla sintetizada en función de los observados en la natural. En la figura 5 aparecen los espectros y los oscilogramas correspondientes a cada segmento. El estímulo sintetizado aparece en la parte superior y el estímulo natural en la parte inferior.

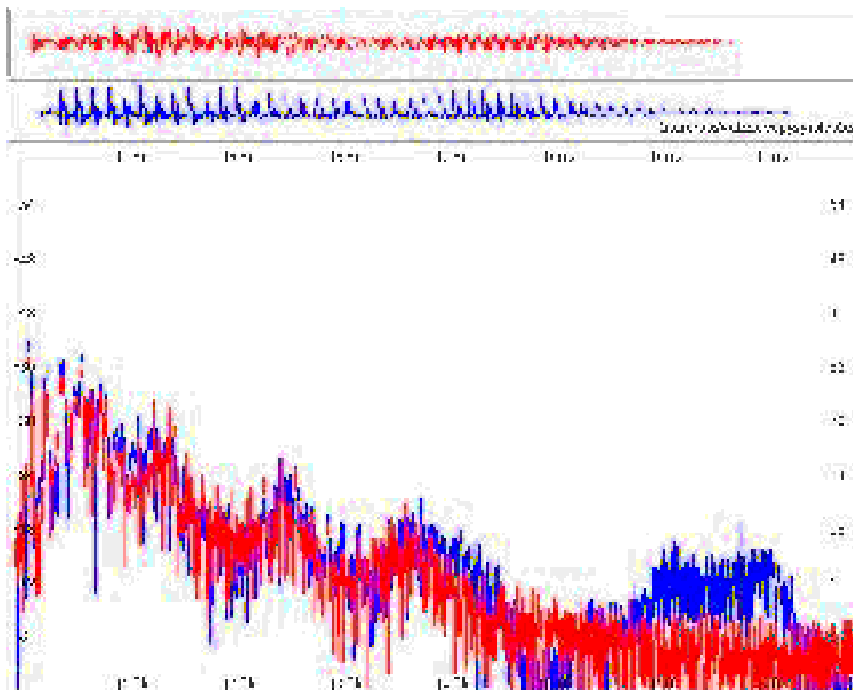


Figura 5. Espectros y oscilogramas de la secuencia [ala] en su versión sintetizada y en su versión natural

A su vez el habla natural puede también modificarse para la preparación de los estímulos que se utilizan en las pruebas de percepción. Las dos técnicas principales son la segmentación, consistente en una modificación en el dominio del tiempo alargando o acortando la duración de un segmento, y el filtrado, mediante el cual se eliminan determinadas bandas de frecuencia.

De forma paralela a lo descrito para los elementos segmentales, una manera de valorar el grado de adecuación de los patrones entonativos es someterlos a una prueba de carácter perceptivo. Los estímulos empleados en la preparación tales pruebas pueden proceder del habla natural (Thorsen, 1979), segmentada o filtrada (Lehiste, 1979), o del habla sintetizada (‘t Hart-Collier-Cohen, 1990; Enríquez *et al.*, 1989, para el español). En el caso del uso de estímulos sintetizados, la técnica LPC permite la modificación de los contornos frecuenciales y su posterior síntesis.

## 3. La conversión de texto a habla

### 3.1. Presentación general

Tal como se ha expuesto al principio, la conversión de texto a habla es una técnica que permite realizar automáticamente la lectura de un texto escrito en ortografía convencional siempre que esté disponible en formato digital. Un sistema de conversión de texto a habla incorpora pues diversos módulos con información fonética y lingüística que realizan una serie de transformaciones a la cadena inicial de caracteres ortográficos hasta convertirla en una onda sonora (Allen *et al.*, 1987; Klatt, 1987) y constituye tanto una aplicación informática útil para acceder mediante el habla a información almacenada en forma escrita como una herramienta de investigación en fonética que permite validar las hipótesis sobre producción y percepción del habla realizadas desde diversos marcos teóricos.

A grandes rasgos, los módulos que configuran un conversor de texto a habla pueden dividirse entre los que realizan un procesado lingüístico y los que se centran en las características acústicas de la onda sonora resultante. Aunque la arquitectura concreta de cada sistema pueda variar, en general el tratamiento lingüístico se realiza en los módulos de pre-tratamiento del texto, transcripción fonética automática, análisis morfológico y sintáctico y en el módulo prosódico, mientras que el procesado acústico se lleva a cabo en un módulo diseñado para tal fin, que incorpora un modelo de síntesis a fin de generar una señal de habla. Algunos ejemplos de sistemas de conversión de texto a habla en español pueden verse en Castejón *et al.* (1994), Martí y Niñerola (1987), Martínez *et al.* (1986), Pérez y Vidal (1991) y Rodríguez *et al.* (1993), entre otros.

En los apartados siguientes se describen algunas estrategias para la incorporación de conocimientos fonéticos y lingüísticos a dos de los módulos de un sistema de conversión de texto a habla: el que se ocupa de la transcripción fonética automática, es decir, del paso de la representación ortográfica a una representación fonética, y el que tiene como objetivo la asignación de elementos prosódicos como la duración, las pausas o la entonación.

### 3.2. La transcripción fonética automática

Un transcriptor o fonetizador es un algoritmo que transforma una cadena de caracteres ortográficos en una cadena de caracteres fonéticos. Un transcriptor, por lo tanto, pone en relación dos representaciones de un mismo texto (la ortográfica y la fonética) a través de una operación informática de transducción. El algoritmo de fonetización - neologismo que utilizamos como sinónimo de "transcripción fonética automática" - ha de llevar a cabo la misma lectura que realizaría un hablante-lector de la lengua, por lo que necesita la información que éste posee. La explicitación del conocimiento del hablante hace de la transcripción fonética automática un campo de investigación lingüística susceptible de múltiples estudios, tal como se describe a continuación.

### 3.2.1. Descripción de la pronunciación

Un sistema de transcripción ha de reflejar de una manera fiel la pronunciación de una lengua; se necesitan, por tanto, descripciones exhaustivas de los fenómenos fonéticos. Aunque las aplicaciones industriales de la transcripción suelen trabajar con la variedad estándar de la lengua, por la ventaja que ofrece de ser un modelo común a todos los hablantes, un fonetizador también puede ser aplicado para transcribir variantes dialectales y sociales si así se requiere: el registro de habla transcrito deberá estar en consonancia con el enunciado que se transcriba. Asimismo, la transcripción es susceptible de ser aplicada a los textos de un determinado periodo histórico de la lengua. En cualquier caso, es necesario partir de estudios que describan la pronunciación de la variante de habla o del periodo elegidos.

### 3.2.2. Descripción de la relación entre la ortografía y la pronunciación

La interpretación fónica de la ortografía que realiza la transducción implica contar con descripciones del valor de los grafemas, de la agrupación silábica de los segmentos y de las normas de acentuación gráfica en las lenguas con acento léxico libre.

Estas descripciones son imprescindibles en la elaboración de un fonetizador ya que la elección de una determinada estrategia de transcripción, por reglas o por diccionario, está condicionada por la complejidad de las normas ortográficas y por la mayor o menor adecuación de éstas a la descripción de las características fonológicas de la lengua. Un sistema de transcripción será más eficaz cuanto más extensa sea su aplicación, con un menor coste, de los medios informáticos empleados y de la información requerida. La transcripción por reglas parece ser el medio más útil en aquellas lenguas que se alejan poco principio de fonémico de la representación ortográfica por el que cada fonema se representa con un único grafema y cada grafema representa un único fonema; por ejemplo, los sintetizadores para el español SINCAS (Martí y Niñerola, 1987) y AMIGO (Rodríguez *et al.* 1993) utilizan transcriptores por reglas. Si la pronunciación de los enunciados no siempre es predecible a partir de la ortografía, se ha de contar con listas de excepciones que corrijan la aplicación de las reglas. Cuanto más se aleje la lengua de aquel principio, más necesaria es la transcripción por diccionario. Como ejemplos de transcriptores que utilizan el léxico podemos citar, para el inglés, el trabajo pionero de Coker *et al.* (1973) y el sistema de fonetización desarrollado en el MIT (Allen *et al.*, 1987).

Es preciso contar con descripciones exhaustivas de lo regular e irregular en la relación ortografía-pronunciación de las lenguas y de los contextos exactos de aplicación de las regularidades e irregularidades. La exhaustividad y la exactitud en la descripción de los fenómenos y de los contextos es una exigencia del medio informático utilizado en la transcripción: toda aquella información necesaria para realizar las operaciones de transducción ha de estar explícita en la codificación del algoritmo.

### **3.2.3. Descripción del conocimiento lingüístico del hablante.**

Una transcripción fonética es la formalización de un determinado conocimiento del hablante, esencialmente, de su conocimiento fonológico: el repertorio de fonemas y de alófonos de una lengua y su distribución; los procesos fonológicos que relacionan alófonos y fonemas; la organización silábica de la cadena fónica y las restricciones de combinación de los grupos consonánticos y vocálicos; la acentuación, incluyendo la determinación del carácter tónico o átono de una determinada unidad léxica, la posición del acento y los fenómenos de reacentuación.

Sin embargo, el conocimiento del hablante que refleja la transcripción trasciende los aspectos meramente fonológicos; por ejemplo, el conocimiento del carácter átono de las partículas forma parte de la competencia fonológica, pero se relaciona con la competencia gramatical en la medida en que se requiere saber la categoría de los elementos léxicos. Es tarea de la investigación lingüística determinar la información que necesita el fonetizador, es decir, el conjunto de conocimientos implicados en el proceso de transcripción.

### **3.2.4. Los estudios que genera la construcción de un transcriptor**

También son fuente de investigación los aspectos derivados de las propias características del sistema automático y de la resolución de problemas técnicos de la formalización.

En la elaboración de un transcriptor se han de tomar determinadas decisiones. En relación con los aspectos teóricos y formales de la transcripción, se ha de decidir qué se transcribe y cómo (Beckman, 1990). Por lo tanto, es imprescindible reflexionar sobre qué debe constituir una transcripción fonética, cuáles han de ser sus objetivos, qué información lingüística ha de aportar y qué convenciones ha de contener un alfabeto fonético para que el sistema de representación sea racional y operativo.

Se ha de determinar a partir de qué tipo de análisis (articulatorio o acústico) se describe la pronunciación de los enunciados. Por ejemplo, en español, la dentalización de /s/ no parece ser relevante acústicamente, según el estudio de Quilis (1966), pero sí articulatoriamente, como señala Martínez Celdrán (1993); una transcripción basada en el estudio articulatorio ha de recoger el alófono dental correspondiente, que no deberá contemplarse si se transcribe tomando como base los resultados acústicos. Además, se deberá decidir el nivel de abstracción de la representación obtenida (transcripción fonológica o transcripción fonética) y la exhaustividad de la descripción: se pueden generar transcripciones fonéticas más o menos estrechas en función de las necesidades de la aplicación y de la investigación. En cada caso, se ha de determinar lo pertinente para la transcripción de cada nivel.

La transcripción fonética, como medio capaz de representar la pronunciación de los enunciados de las lenguas ha de poseer un mecanismo formal claro, que responda a unos objetivos lingüísticos concretos. Los inventarios fonéticos que se conocen en la actualidad -por ejemplo, el Alfabeto Fonético Internacional (AFI) de la *International*

*Phonetic Association* (IPA) y el alfabeto de la *Revista de Filología Española* (RFE) - presentan numerosos problemas: las convenciones permiten la posibilidad de una doble representación de los sonidos y los diacríticos adoptados no contemplan todas las caracterizaciones fonéticas (De la Mota y Ríos, 1993). A éstos problemas se suman las limitaciones del medio informático, allí donde la técnica condiciona los medios de representación lingüística: la proliferación de diacríticos complica la codificación informática del alfabeto, como puede observarse en las propuestas de Esling (1988) y Esling y Gaylor (1993) para el AFI; no todos los sistemas operativos (por ejemplo, el VAX-VMS) poseen fuentes fonéticas y la configuración de los teclados de los ordenadores no representa todos los símbolos y diacríticos de los alfabetos fonéticos comúnmente utilizados en el ámbito de la lingüística; no siempre los alfabetos fonéticos utilizados en los medios informáticos poseen representación para el repertorio de alófonos que se haya fijado la transcripción, lo cual ha llevado a propuestas como la de Wells (1987, 1990) para el desarrollo del proyecto SAM (*Multilingual Speech Input/Output Assessment, Methodology and Standardisation*) basadas en un enfoque fonológico, posteriormente complementadas con codificaciones del AFI de carácter más fonético (Wells, 1995).

### **3.2.5. Transcripción fonética automática y lingüística**

El carácter utilitario de todo transcriptor no significa que se desvincule necesariamente de los estudios lingüísticos teóricos. Un sistema puede implementar las propuestas de un modelo fonológico, como hacen Howard y Goldman (1994) en su transcriptor del español, donde las reglas de silabificación siguen el algoritmo de Hualde (1991), inscrito en la fonología generativa.

El diseño mismo de un sistema de transcripción fonética automática puede responder a los principios de una determinada teoría lingüística. El algoritmo creado por Laporte (1988) para la generación del DELAP (*Dictionnaire Electronique du Laboratoire d'Automatique Documentaire et Linguistique pour la Phonemique*) se inscribe dentro de los presupuestos de la gramática léxica, elaborados por Gross (1975) a partir de los estudios de Zellig Harris. El objetivo de una gramática léxica es realizar "una descripción lingüística sistemática", que ha de entenderse como la enumeración estructurada de las reglas gramaticales que definen una lengua, así como la representación exacta de las unidades del léxico en las que se aplican. Para generar el DELAP, se parte de un diccionario electrónico ortográfico del francés en el que se codifican numéricamente, con una particularidad de lectura, aquellos elementos léxicos cuya pronunciación no se puede deducir a partir de la ortografía (existen 200 tipos distintos de particularidades de lectura). El algoritmo de fonetización calcula la pronunciación de los elementos léxicos a partir de dos informaciones: las reglas generales y el tipo de particularidad de lectura que se aplica a cada palabra. Además, el DELAP está concebido para generar la representación fonética de las formas flexivas de nombres, adjetivos y verbos a partir de la representación ortográfica canónica de las palabras: todas las entradas léxicas están codificadas alfanuméricamente para indicar la pronunciación de su paradigma flexivo, por lo que las reglas morfológicas vinculadas con la flexión están representadas en las unidades del léxico en las que se aplican.

Un transcriptor, como instrumento, puede beneficiarse de la aplicación de los métodos y conocimientos de la lingüística. El procedimiento utilizado para el tratamiento de las irregularidades en la creación del DEFE (Diccionario Electrónico Fonético del Español), en el *Laboratori de Lingüística Informàtica* del Departament de Filologia Espanyola de la Universitat Autònoma de Barcelona, aplica principios generales de la lengua al tratamiento de las irregularidades ortográficas, en palabras con secuencias consonánticas anómalas en español (Ríos, 1993):

- La aplicación de las reglas de silabificación generan secuencias regulares, por ejemplo: entre dos consonantes que no formen grupo tautosilábico siempre habrá un límite silábico: *pa-lim-p-sesto* / *an-g-s-trom*.

- Después de la silabificación se aplica una regla que recoge un principio general de la sílaba en español: "ningún segmento consonántico puede formar sílaba aislado, sin apoyarse en una vocal". Esta regla borra cualquier consonante situada entre dos límites silábicos consecutivos y se completa con otra regla que resilabifica [s], única consonante que forma grupo consonántico en la rima: *ans-trom*.

Si lo comparamos con la tradicional lista de excepciones o con la multiplicación de reglas *ad-hoc*, una para cada contexto, como hacen Cabrera *et al.* (1991), este procedimiento, al basarse en principios fonológicos generales de la lengua, posee mayor capacidad descriptiva: se aplica a todas las palabras que tengan el mismo esquema ortográfico, y mayor capacidad explicativa: se puede aplicar a cualquier palabra de nueva incorporación; además, resulta informáticamente más económico.

En estas líneas sólo hemos esbozado algunos aspectos concernientes a la fonetización como materia de investigación lingüística. A modo de conclusión: la transcripción fonética del habla debe constituir un método eficaz para representar mediante símbolos la pronunciación de enunciados de cualquier lengua; el desarrollo de transcriptores automáticos capaces de cumplir esa tarea con un margen de error virtualmente nulo es una materia de investigación de gran interés para la lingüística, por sus aplicaciones tecnológicas (procesamiento lingüístico para la conversión de texto a habla y en el almacenamiento de bases de datos lingüísticas, enseñanza de lenguas, corrección ortográfica, etc.) y por ser fuente de estudios que redundan en un mayor conocimiento de las lenguas.

### 3.3. Elementos para un modelo prosódico en la conversión de texto a habla

Uno de los módulos que mayor interés despierta actualmente en el proceso de conversión de texto a habla es precisamente el que se ocupa determinar las características suprasegmentales del habla, ya que de ellas depende en buena parte la naturalidad de la salida vocal de un sistema de síntesis. Por ello, presentamos en los siguientes apartados algunos elementos que configuran este módulo como la asignación y modelización de la duración segmental, la inserción automática de pausas y la definición de un modelo de entonación.

### 3.3.1. La duración segmental

La complejidad teórica que entraña el análisis de la duración segmental proviene, básicamente, del elevado número de factores que convergen en este fenómeno lingüístico. Probablemente sea ésta su característica más destacable. Por ello, la elaboración de un modelo de duración debe hacer especial hincapié en el adecuado tratamiento de estos factores y de las relaciones que mantienen entre sí.

La organización temporal del enunciado es resultado de la interacción de fenómenos diversos, como su extensión, la duración intrínseca propia de cada segmento, el número de alófonos de la sílaba, el acento léxico, la estructura sintáctica o la información que tal enunciado aporta (Di Cristo, 1985; Ríos, 1991 y Marín, 1995 para el español). El acento, la posición en la frase, la estructura silábica, la posición en la sílaba, el carácter sordo o sonoro de los sonidos adyacentes o la posición respecto a la sílaba acentuada son algunos de los factores que aparecen de forma más recurrente en los trabajos dedicados a la duración segmental.

No obstante, en la revisión que presentamos aquí, trataremos únicamente, por motivos de espacio, la influencia del acento y la posición en la frase, además de la duración intrínseca de cada sonido.

Uno de los problemas teóricos que debemos plantearnos a la hora de desarrollar un modelo de duración no es otro que el de hallar la unidad lingüística en la que se estructura la información temporal. Este aspecto resulta especialmente controvertido, ya que las unidades propuestas son muy diversas: la sílaba, la palabra, el grupo acentual, el grupo fónico, etc. (Noteboom, 1991; Fant, 1991). Por ello, en el presente apartado dejaremos de lado esta cuestión.

#### 3.3.1.1. La duración vocálica

Al estudiar la duración de los sonidos vocálicos en español, podemos constatar, en primer lugar, que las vocales poseen una duración propia y característica: en un mismo contexto, cada una de las vocales se manifiesta sistemáticamente con diferentes valores duracionales con respecto al resto de vocales.

Sobre este punto concreto, Marín (1995) considera que podemos dividir las vocales en tres grupos diferenciados: [i] y [u] son las vocales que presentan una menor duración, seguidas de [e] y [o] y, por último, de [a]. Estos resultados son muy similares a los que aparecen en Navarro Tomás (1916). Como vemos, la duración intrínseca de las vocales se relaciona con un parámetro articulatorio: una mayor obertura se corresponde con una mayor duración.

Por lo que respecta al acento, cabe indicar que una vocal acentuada presenta una mayor duración que la misma vocal no acentuada en un contexto idéntico. Tanto Borzone y Signorini (1983) como Marín (1995) coinciden en señalar este comportamiento.

La posición en la frase es una de las variables que más claramente incide en la cantidad de los sonidos: la duración de una vocal aumenta considerablemente cuando se encuentra en posición prepausal. Sobre este fenómeno, que se conoce con el nombre de *Prepausal Lengthening Effect*, existe un amplio consenso en los trabajos dedicados a la duración segmental. Varios autores -Navarro Tomás (1916), Borzone y Signorini

(1983), Santos *et al.* (1988), Macarrón *et al.* (1991) y Marín (1995), entre otros-constatan la existencia de este fenómeno.

En el caso de la duración vocálica, el análisis de las vocales en contacto y de los diptongos necesita de un análisis específico, tal como puede observarse en Aguilar (1991).

### 3.3.1.2. La duración consonántica

Al igual que las vocales, las consonantes también poseen una duración intrínseca. A este respecto, del Barrio y Torner (1995) plantean que las consonantes del español se pueden ordenar, de mayor a menor duración, del siguiente modo: consonantes sordas, vibrante múltiple, nasales, fricativas sonoras, laterales y vibrante simple.

Como se puede observar, en lo que a su duración se refiere, la agrupación de las consonantes puede realizarse según el carácter sordo o sonoro y el modo de articulación.

No existe un claro consenso sobre la influencia del acento en la duración consonántica. Así, mientras que Borzone y Signorini (1983) afirman que las consonantes que se encuentran en sílaba acentuada presentan una mayor duración que las que aparecen en sílaba no acentuada, del Barrio y Torner (1994) señalan que el acento únicamente produce el alargamiento de [n], [l] y [r] cuando, además, se encuentran en coda silábica. Los resultados que aparecen en Ríos (1991) y en Iglesias (1994) también parecen indicar que el acento no es un factor que tenga una clara incidencia en la duración de las consonantes del español.

Varios son los autores que coinciden en señalar un alargamiento de la duración de las consonantes cuando se encuentran en posición prepausal. Así lo afirman Navarro Tomás (1918), Borzone y Signorini (1983) y del Barrio y Torner (1994), entre otros. Navarro Tomás (1918) resalta, además, el mayor alargamiento de las laterales en este contexto.

### 3.3.1.3. La modelización de la duración

A la hora de implementar los resultados de un análisis de duración segmental en un conversor de texto a habla, disponemos de varios sistemas de representación y cálculo de la duración.

Así, por ejemplo, podemos elaborar una base de datos en la que aparezca un valor de duración específico para cada sonido en un contexto determinado, desarrollar sistemas de reglas o ecuaciones, o representar nuestros datos mediante árboles binarios. Las diferencias entre estos sistemas de representación y cómputo de la duración son importantes, ya que la elección de uno u otro tipo puede obligarnos a estructurar nuestros datos de tal forma que se establezcan reglas más generales y de mayor alcance.

La modelización de los datos duracionales no consiste únicamente en buscar el método más adecuado para representar y calcular dichos datos. En este proceso aparecen también algunos problemas teóricos de gran interés como, por ejemplo, la influencia simultánea de dos o más factores. Disponemos de tres posibilidades básicas, presentadas detalladamente en van Santen y Olive (1990): el efecto combinado de dos o

más factores es mayor que la suma de las influencias de cada uno de los factores por separado (interacción), es menor (incompresibilidad) o igual (juntura independiente).

El problema que acabamos de plantear debe ser tenido en cuenta en el momento de realizar el diseño experimental. En este caso, pensamos que una posición posicionamiento apriorística queda fuera de lugar. Serán los datos experimentales los que nos indicarán cuál de las tres posibilidades afecta a la relación entre dos factores concretos.

Otro de los aspectos que emergen a la hora de modelizar la duración es el de su carácter absoluto o relativo; esto es, si la influencia de un factor (p.ej. el acento) en un sonido determinado es absoluta (debemos añadir una cantidad en milisegundos) o es relativa a la duración que ya posee ese sonido (debemos añadir un tanto por ciento). En el caso del español, tanto Macarrón *et al.* (1991) como Marín (1994) proponen la utilización de modelos multiplicativos.

Finalmente, cabe señalar la existencia de otras aproximaciones al problema de la modelización de la duración como, por ejemplo, los árboles de regresión de Riley (1992) o la propuesta de Campbell (1992).

### **3.3.2. La asignación automática de pausas**

La complejidad a la hora de modelar de forma adecuada la duración segmental descrita en el apartado anterior se reproduce en el caso de la asignación automática de pausas. Por un lado, factores de diferente naturaleza condicionan la segmentación prosódica de un texto realizada por un hablante: desde factores fisiológicos, como la necesidad de respirar, hasta factores sociolingüísticos, psicolingüísticos o lingüísticos (Goldman-Eisler, 1961; Dechert y Raupach, 1980; Nespor y Vogel, 1983; Cruttenden, 1986); por otro lado, ciertas pausas son opcionales, lo que dificulta cualquier intento de sistematización de este fenómeno.

Cabe destacar, además, que si el modelo de predicción de la aparición y localización de pausas se incorpora en un sistema de conversión de texto a habla, se requiere, a la vez, que pueda ejecutarse en tiempo real y que analice cualquier tipo de texto.

Sin embargo, a pesar de estas dificultades metodológicas, el interés en el desarrollo de procedimientos automáticos de asignación de pausas es creciente, debido principalmente al desarrollo de sistemas multilingües en prototipos de laboratorio y al crecimiento de los servicios vocales que usan la síntesis de habla. Por otro lado, se intenta solventar así uno de los principales problemas en los sistemas de conversión de texto a habla: la falta de naturalidad debida a las deficiencias en la información prosódica y lingüística señalada al principio (Boëffard *et al.*, 1996).

En general, se reconoce que una adecuada segmentación del texto en grupos fónicos contribuye a la mejora de la inteligibilidad y de la aceptabilidad de este tipo de sistemas por parte del usuario. Ello es así debido al papel que desempeña la prosodia en la percepción del habla: en concreto, las pausas ayudan a segmentar en palabras y en grupos de palabras, además de ofrecer un tiempo adicional para procesar la información lingüística (Nooteboom *et al.*, 1978; Scharpff y van Heuven, 1988).

Con el principal objetivo de contribuir a la mejora de la calidad de los sistemas de conversión de texto a habla, se han desarrollado una serie de algoritmos de segmentación prosódica que abordan el problema desde perspectivas diferentes: desde partir de un análisis sintáctico completo que posteriormente es reinterpretado en términos prosódicos (Frenkenberger *et al.*, 1994) hasta prescindir de la información sintáctica para centrarse únicamente en cuestiones morfológicas (Emerard *et al.*, 1992).

Para el español, por ejemplo, encontramos en el sistema de conversión de texto a habla de Telefónica I+D un módulo de asignación automática de pausas basado en las categorías morfológicas de las palabras (Castejón *et al.*, 1994). El módulo examina la secuencia de categorías de la frase en dos etapas: los signos de puntuación - considerados como una categoría más - se corresponden invariablemente con una pausa; si la secuencia resultante es demasiado larga, se asignan unas pausas adicionales de acuerdo con criterios morfológicos: ciertas categorías gramaticales favorecen más que otras la aparición de una pausa. Dichas categorías, en orden de prioridad, son: las conjunciones coordinantes, las conjunciones subordinantes, los verbos y las palabras funcionales. El módulo examina la secuencia y asigna la pausa en el lugar correspondiente.

Por su parte, López (1993) desarrolla un modelo de asignación prosódica basado en unos coeficientes - del 1 al 9 - que indican la relación más o menos estrecha entre las palabras de una frase: cuanto más alto es el coeficiente, mayor probabilidad de aparición de pausa, dado que se interpreta que el grado de cohesión entre esas dos palabras contiguas es bajo. La asignación de pausa depende de ese coeficiente de relación, del cual es responsable la categoría gramatical de la palabra.

Sin embargo, el enfoque considerado actualmente como más adecuado es el que conjuga en la predicción de las fronteras prosódicas la importancia de los factores sintácticos y no sintácticos (Ladd, 1987; Bachenko y Fitzpatrick, 1990; Monaghan, 1992; Quené y Kager, 1992; Hirschberg y Prieto, 1994; Gili y Quazza, 1996). Los algoritmos desarrollados en esta línea comparten la idea de que los fenómenos suprasegmentales no pueden derivarse únicamente de la estructura sintáctica, sino que hay que acudir además a informaciones como la longitud de la frase o la función discursiva de las palabras.

Desde este punto de vista, no se considera necesario un análisis sintáctico exhaustivo. Por ejemplo, las reglas propuestas por Bachenko y Fitzpatrick (1990) sólo utilizan un subconjunto de la información sintáctica ofrecida por un analizador. Dichas reglas han de tener acceso a la categoría léxica, a los núcleos, a la distribución y al orden de constituyentes, pero no necesitan conocer las relaciones entre predicados y argumentos, la distribución en cláusulas o la posición de los modificadores en la frase.

O'Shaughnessy (1990), por su parte, considera que para generar de forma adecuada la distribución de las pausas es suficiente conocer la posición de los verbos en la frase, las fronteras sintácticas mayores y las palabras acentuadas.

En el caso del español, Casacuberta *et al.* (1996), implementando los resultados obtenidos en el análisis experimental presentado en Marín *et al.* (1996), proponen un tratamiento en el que se combina la información sintáctica y la prosódica. La unidad

básica de dicho trabajo es el grupo acentual categorizado (GAC), que podría definirse como un grupo acentual (GA) etiquetado sintácticamente.

El modelo de asignación de pausas consta de las siguientes fases: 1) segmentación del texto en GAs; 2) etiquetado sintáctico de los GAs; 3) aplicación de un conjunto de restricciones prosódicas, entre las cuales cabe destacar la longitud de la frase (en número de GAs) y la distancia mínima (también en número de GAs) que debe existir entre una pausa y el inicio o el final de un enunciado; 4) aplicación de una jerarquía, basada en criterios sintácticos, sobre la probabilidad de aparición de pausa delante de un GAC. Al final de este proceso, se determina si una pausa es obligatoria o no, y en caso afirmativo, el lugar adecuado para su aparición.

Por último, el uso de dominios fonológicos permite disponer de una unidad entre la palabra y el constituyente sintáctico que resulta útil para la predicción de la distribución de las pausas en las frases. Hirschberg y Prieto (1994) adoptan la teoría entonativa de Pierrehumbert (Pierrehumbert, 1980; Beckman y Pierrehumbert, 1986) para desarrollar un módulo de segmentación prosódica integrado en un conversor de texto a habla del español mexicano. Las reglas se adquieren de forma automática a partir de texto anotado, mediante la técnica estadística de obtención de datos y de parametrización de las variables descrita en Riley (1988) -*CART, Classification And Regression Trees* -. Esta técnica también se empleó para el desarrollo de un modelo similar en un conversor del inglés (Wang y Hirschberg, 1991).

A modo de conclusión, cabe mencionar que en el problema de la asignación automática de pausas se pone de manifiesto la necesidad de una relación entre disciplinas como la prosodia, la sintaxis y la tecnología del habla, con el fin de establecer un modelo en el cual se definan las unidades y los niveles de análisis adecuados. Por otro lado, desde el punto de vista de las aplicaciones, es necesario el desarrollo de sistemas con un bajo coste computacional y que no impongan restricciones sobre los textos de entrada.

### 3.3.3. Modelos entonativos

Los sistemas de conversión de texto a habla incluyen como parte de su módulo prosódico un submódulo entonativo. Este módulo utiliza normalmente la información proporcionada por los módulos de análisis lingüístico del conversor para determinar la asignación de una cadena de valores de  $F_0$  (frecuencia fundamental) asociada al enunciado que va a ser sintetizado. La descripción de las curvas melódicas de una lengua que subyace a estos módulos recibe habitualmente el nombre de modelo entonativo, entendiéndose por entonación el equivalente a “curva melódica” o “evolución de la frecuencia fundamental a lo largo del tiempo”.

El desarrollo de modelos entonativos ha sido una tarea abordada en los últimos años tanto por ingenieros como por lingüistas, usando herramientas y enfoques teóricos muy diferentes. Se han propuesto modelos para lenguas como el alemán, el danés, el francés, el inglés americano, el inglés británico, el japonés, o el sueco, entre otras, tantos que resulta imposible enumerarlos todos en este breve espacio. Para el español, se han publicado algunos trabajos en esta dirección (Fant, 1984; Garrido, 1991; Fujisaki *et al.*, 1994; López *et al.*, 1994; de la Mota, 1995; Garrido, 1996).

Con independencia del enfoque adoptado, el desarrollo de un modelo entonativo implica normalmente una primera fase de representación de las curvas melódicas, que facilite el análisis posterior, y una segunda fase de descripción de los patrones melódicos y de determinación de las reglas que controlan su uso, que implica muchas veces una validación perceptiva de los mismos.

Durante la fase de representación, las curvas melódicas se convierten en una cadena de símbolos convencionales (transcripción) o en una representación simplificada de la forma de la curva (estilización). El uso de un sistema de transcripción se asocia normalmente a aproximaciones fonológicas (de niveles), en tanto que la estilización se ha asociado más a aproximaciones fonéticas (por contornos). Suele evitarse la incidencia de la micromelodía tomando en consideración sólo la información frecuencial procedente de los núcleos silábicos. La fase de estilización puede ir seguida de un proceso de normalización. Los valores frecuenciales se pueden normalizar a partir de la caracterización de la frecuencia fundamental habitual de cada hablante, ya sea a partir del rango -o diferencia entre el máximo y mínimo-, como en el modelo de Takefuta (1975), o a partir del valor mínimo -con el diseño previo de una línea de base-, como en el caso del sistema propuesto por Pierrehumbert (1980). Otra manera de normalizar los datos de frecuencia fundamental es estudiar las diferencias entre los diversos puntos que definen la evolución del contorno frecuencial, más que el valor de los puntos en sí. Tanto la formulación matemática del análisis por semitonos (logarítmico) de 't Hart, Collier y Cohen (1990) otros análisis de naturaleza lineal determinan la distancia relativa existente entre dos valores de frecuencia. En de la Mota (1995) se propone un sistema basado en el cálculo de desniveles frecuenciales, de tal manera que se normalizan a la vez la información temporal y la frecuencial.

En la figura 6 se presenta un ejemplo de curva estilizada obtenida automáticamente, en la que la curva original se ha reducido a una serie de puntos de inflexión unidos por líneas rectas. Esta representación ha sido obtenida por medio del sistema descrito en Jiménez (1994).

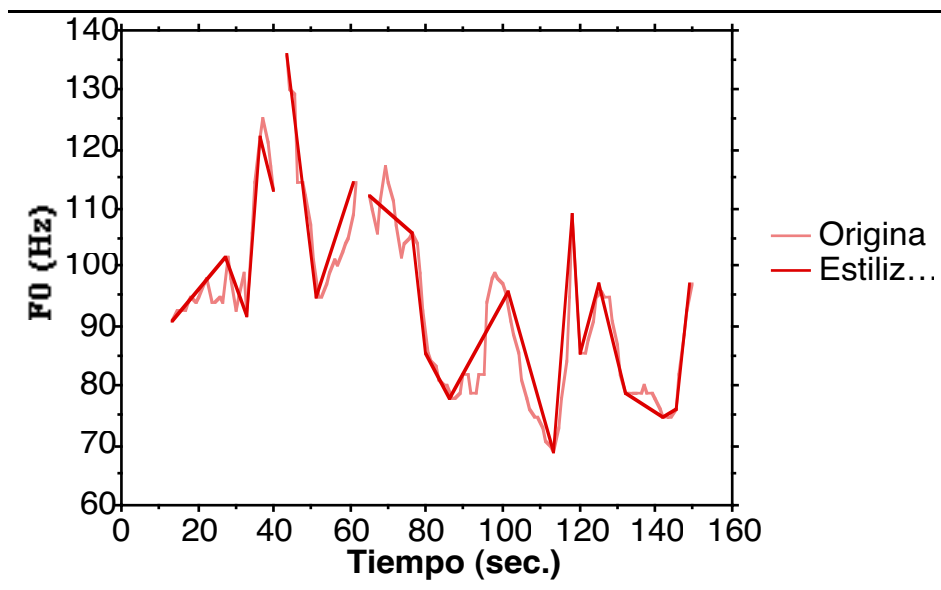


Figura 6. Curva melódica original (en línea discontinua) y curva estilizada correspondiente al enunciado 'La organización terrorista ha protagonizado en lo que va de año' pronunciado por un locutor masculino.

En la fase de descripción, se realiza normalmente la estandarización o definición de los patrones entonativos. En unos casos, éstos se definen también utilizando procedimientos automáticos o semi-automáticos, o con métodos estadísticos (ten Bosch, 1993). En otros, la definición se lleva a cabo manualmente (Garrido, 1996). El objetivo es, en cualquier caso, obtener una descripción formal de los contornos típicos de entonación y definir explícitamente los factores que determinan su uso.

Los modelos entonativos se pueden dividir, siguiendo la clasificación propuesta por Ladd (1988), en lineales y jerárquicos. Los modelos lineales consideran que la entonación es el resultado de la concatenación lineal de diferentes niveles tonales - por ejemplo, *H(igh)*, *L(ow)* -, y que éstos se construyen de izquierda a derecha, en un solo ciclo (Pierrehumbert, 1980). Los modelos jerárquicos, en cambio, suponen que los contornos entonativos se generan en varios ciclos ('t Hart *et al.*, 1990). En el caso de los modelos jerárquicos, se supone la existencia de patrones de diferentes niveles, que se superponen para la obtención del contorno final. El número de niveles considerados varía de unos modelos a otros, pero como mínimo, dos niveles están siempre presentes: un nivel global, que define la forma general de la curva melódica a lo largo del enunciado, y un nivel local, que modela los picos y valles de la curva.

La forma de los patrones globales se ha asociado típicamente en español a la modalidad oracional (Navarro Tomás, 1944; Garrido, 1991). Sin embargo, también parece estar relacionada con otros factores, como la duración del enunciado. Estos patrones se han asociado típicamente al ámbito de la unidad melódica, aunque también se ha observado que pueden definirse patrones globales de ámbito superior, que abarcan incluso un párrafo entero (Garrido *et al.*, 1993).

Una manera de representar los patrones globales es mediante una línea que determina la evolución general de la curva melódica a lo largo de un enunciado; sobre esta línea se superpondrán las variaciones locales debidas al acento. Esta es la aproximación

subyacente en descripciones clásicas del español como la de Navarro Tomás (1944) y en los patrones propuestos en Garrido (1991). Sin embargo, la evolución global de la curva melódica también puede representarse de otras formas, como por ejemplo por medio de líneas, paralelas o convergentes, que definen los diferentes niveles que puede alcanzar la curva entonativa en cada instante de tiempo. Ésta es la aproximación seguida, por ejemplo, en 't Hart *et al.* (1990), y para el español, en Garrido *et al.* (1993) y Garrido (1996). El tipo de representación obtenido se ejemplifica en la figura 7.

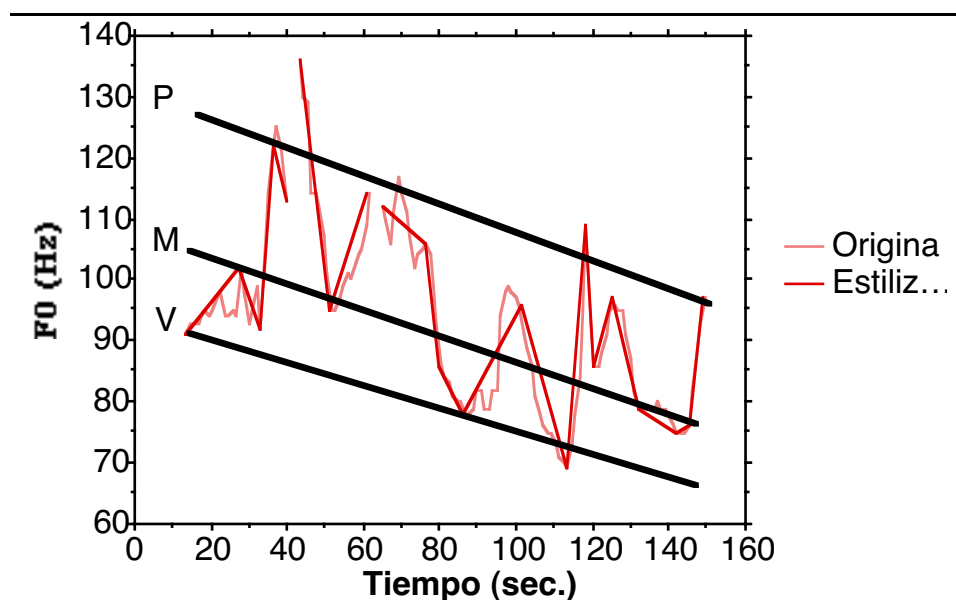


Figura 7. Contorno entonativo (original y estilizado) correspondiente al enunciado 'La organización terrorista ha protagonizado en lo que va de año', pronunciado por un locutor masculino. El patrón melódico global está modelado por medio de tres líneas descendentes que definen tres niveles teóricos a lo largo de la curva.

En el caso de los patrones locales, se suele distinguir entre los patrones en posición inicial, interior y final de grupo melódico. Los patrones en posición final son los que se asocian a los llamados tonemas (Navarro, 1944) o junturas terminales (Quilis, 1993) descritos en la bibliografía. La forma de los patrones iniciales e interiores se ha relacionado típicamente con la localización del acento léxico o del enfático (de la Mota, 1995), aunque últimamente también se ha analizado su relación con la estructura sintáctica del enunciado (Llisterri *et al.*, 1995). Por lo que se refiere a los patrones finales, se han estudiado tradicionalmente en relación con la modalidad oracional o el límite sintáctico ante el que aparecen (Estruch y Garrido, 1995). El ámbito de estos patrones no está aún muy definido, aunque suelen asociarse con sílabas, grupos acentuales o grupos tónicos.

Los patrones locales se pueden definir, en función del tipo de representación (transcripción o estilización) como series de niveles asociados a determinados puntos del enunciado (Pierrehumbert, 1980, por ejemplo) o como series de movimientos que definen un contorno típico ('t Hart *et al.*, 1990). Un ejemplo de modelización por contornos de los patrones locales de un enunciado se ofrece en la figura 8.

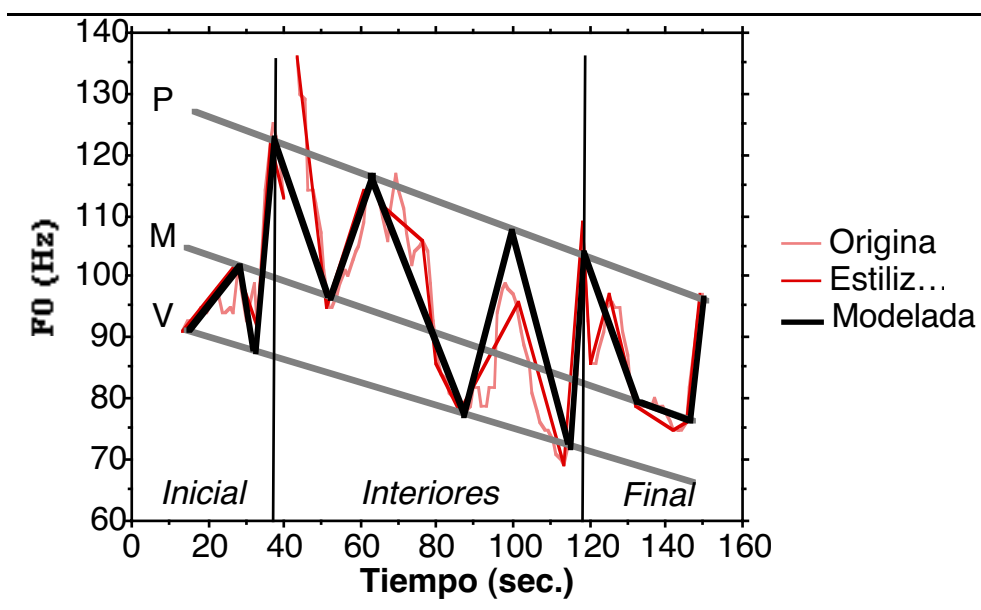


Figura ·8. Contorno entonativo (original, estilizado y modelado) correspondiente al enunciado ‘La organización terrorista ha protagonizado en lo que va de año’, pronunciado por un locutor masculino. Los diferentes patrones locales (inicial, interiores y final) se han superpuesto al patrón global para formar la curva modelada.

Por último, los patrones obtenidos deben ser validados perceptivamente para comprobar la interpretación que éstos reciben por parte de los receptores. La validación puede hacerse directamente mediante la implementación en el sistema, o bien mediante la realización de pruebas de percepción. En este caso, es necesario disponer de una herramienta adecuada que permita la modificación de las curvas melódicas y la síntesis con el contorno modificado.

## 4. Conclusión

A lo largo de este capítulo se ha pretendido ofrecer una muestra de cómo el surgimiento de nuevas herramientas informáticas incide en la metodología de una disciplina lingüística como es la fonética y, a la vez, permite incorporar los conocimientos obtenidos a sistemas de utilidad práctica como los conversores de texto a habla. Con ello se ha querido poner de manifiesto, tomando como punto de partida la experiencia de nuestro grupo de investigación, la imbricación existente entre los métodos de análisis del habla y los resultados que pueden obtenerse en la descripción fonética de las lenguas, haciendo patente también que una aproximación experimental a la fonética lleva al desarrollo de sistemas de comunicación persona-máquina en el área de las tecnologías del habla. Esta doble perspectiva es posible partiendo de la base de una teoría que oriente la utilización de los métodos y herramientas de observación sin negligir los requisitos prácticos impuestos por la necesidad de crear aplicaciones. La informática juega en este proceso un papel relevante, puesto que constituye simultáneamente una herramienta y un entorno de desarrollo de sistemas, los cuales, a su vez, plantean problemas que deben solucionarse con nuevos métodos contribuyendo, en última instancia, a una renovación de las teorías.

## 5. Referencias

- ABBS, J.H. - K.L. WATKIN (1976) "Instrumentation for the Study of Speech Physiology", en N.J. LASS (Ed.) *Contemporary Issues in Experimental Phonetics*. New York: Academic Press. pp. 41-75.
- ABERCROMBIE, D. (1967) *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- AGUILAR, L. (1991) *Algunas cuestiones en torno a la reducción fonética en secuencias de vocales en contacto*. Trabajo de investigación de tercer ciclo. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.
- AGUILAR, L. (1994) *Los procesos fonológicos y su manifestación fonética en diferentes situaciones comunicativas: la alternancia vocal/ semiconsonante/ consonante*. Tesis doctoral. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.
- AGUILAR, L.-B. BLECUA -M. MACHUCA -R. MARÍN (1993) "Phonetic Reduction Processes in Spontaneous Speech", en *Eurospeech'93. 3rd European Conference on Speech Communication and Technology. Berlin. Germany. 21-23 September 1993*. Vol. 1. pp. 433-436.
- AGUILAR, L.-M. ANDREU (1991) "Acoustic description of Spanish approximants in laboratory speech and in continuous speech", en *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications. Vol 3. pp. 362-365
- ÁLVAREZ HENAO, L.E. (1977) *Fonética y Fonología del Español*. Colombia: Ediciones La Catedral.
- ALLEN, J.- HUNNICUTT, M.S.- KLATT, D.H. (with R.C. ARMSTRONG and D. PISONI) (1987) *From Text to Speech: The MITalk System*. Cambridge: Cambridge University Press (Cambridge Studies in Speech Science and Communication).
- ATAL, B.S. (1985) "Linear Predictive Coding of Speech", en FALLSIDE, F.- WOODS, W.A. (Eds.) *Computer Speech Processing*. Englewood Cliffs, N.J.: Prentice Hall International. pp. 81-124.
- ATAL, B.S. - HANAUER, S.L. (1971) "Speech analysis and synthesis by linear predictive coding of the speech wave", *Journal of the Acoustical Society of America* 50: 637-55; en J.L. FLANAGAN- L.R. RABINER (Eds.) (1973) *Speech synthesis*. Stroudsburg, Penn.: Dowden, Hutchinson & Ross. pp. 270-278.
- BACHENKO, J. - FITZPATRICK, E. (1990). "A computational grammar of discourse-neutral prosodic phrasing in English", *Computational Linguistics* 16, 3: 155-170.
- BALL, M.J. (1989) *Phonetics for speech pathology*. London - New Jersey: Whurr Publishers Ltd.
- BARRIO, L. del - TORNER, S. (1995) "La duración consonántica en castellano", Comunicación presentada en el XXV Simposio de la Sociedad Española de Lingüística, Zaragoza, 11-14 de diciembre de 1995. Resumen publicado en: *Revista Española de Lingüística* 26,1: 126-127.
- BECKMAN M. E. (Ed.) (1990) *Phonetic Representation. Journal of Phonetics* 18, 3.
- BECKMAN, M.- PIERREHUMBERT, J. (1986) "Intonational structure in Japanese and English", *Phonology Yearbook*, 3: 15-70.
- BECKMAN, M.E. (1988) "Phonetic Theory", en NEWMAYER, F.J. (Ed.) *Linguistics: The Cambridge Survey. Vol I. Linguistic Theory: Foundations*. Cambridge: Cambridge University Press. pp. 216-238. Trad. cast. de L. A. Santos: "Teoría Fonética", en NEWMAYER, F.J. (Ed) *Panorama de la lingüística moderna de la Universidad de Cambridge. I Teoría lingüística: Fundamentos*. Madrid: Visor (Lingüística y Conocimiento, 7) 1990. pp. 259-282.
- BLECUA, B. (1996) *Caracterización acústica de las vibrantes del español en posición intervocálica*. Trabajo de investigación de tercer ciclo. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.
- BOËFFARD, O. (1993) *Segmentation automatique d'unités acoustiques pour la synthèse de la parole*. Tesis Doctoral. Université de Rennes I- CNET.
- LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autònoma de Barcelona - Editorial Milenio. pp. 449-479.

BOËFFARD, O. *et al.* (1996) "Utilisation de techniques d'apprentissage automatique pour les traitements linguistiques et prosodiques en synthèse de la parole: quelques résultats en Anglais, Allemand et Français", *Actes des XXIèmes Journées d'Études sur la Parole, Avignon, France*. pp. 383-386.

BORZONE DE MANRIQUE, A.M. - SIGNORINI, A. (1983) "Segmental Duration and Rythm in Spanish", *Journal of Phonetics* 11: 117-128.

BOSCH, L. ten (1993) "Algorithmic classification of pitch movements", *Working Papers. Lund University, Department of Linguistics* 41 (*Proceedings of an ESCA Workshop on Prosody, September 27-29, 1993, Lund, Sweden*), pp. 242-245.

CABRERA, C. - CONTINI M. - BOË L.-J. (1991) "La phonétisation du castillan", *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications. Vol 4. pp. 114-117.

CAMPBELL, W.N. (1992) "Syllable-based segmental duration", en BAILLY, G.- BENOÏT, C. (Eds.) *Talking Machines. Theories, Models and Designs*. Amsterdam: North-Holland / Elsevier Science Publishers. pp. 211-224

CASACUBERTA, D.- MARÍN, R.- AGUILAR, L. (1996) "A formal description of a syntactico-prosodic analysis for unrestricted text", en *Proceedings of the II International Conference on Mathematical Linguistics, Universitat Rovira i Virgili, Tarragona*

CASTEJÓN, F.- ESCALADA, G.- MONZÓN, L.- RODRÍGUEZ, M.A.- SANZ, P. (1994) "Un conversor texto-voz para el español", *Comunicaciones de Telefónica I+D* 5, 2: 114-131.

CODE, C.- BALL, M. (Eds.) (1984) *Experimental Clinical Phonetics. Investigatory Techniques in Speech Pathology and Therapeutics*. London & Camberra: Croom Helm.

COKER, C. H. - UMEDA, N. - BROWMAN, C. P. (1973) "Automatic Synthesis from Ordinary English Text", en J.L. FLANAGAN - L.R. RABINER (Eds.) *Speech synthesiss*. Stroudsburg, Penn.: Dowden, Hutchinson & Ross. pp. 400-411.

COLE, R.A.- MARIANI, J.- USZKOREIT, H.- ZAENEN, A.- ZUE, V. (Eds.) (1996) *Survey of the State of the Art in Human Language Technology*. Publicación electrónica. URL: <http://www.cse.ogi.edu/CSLU/HLTsurvey/HLTsurvey.html>

CRUTTENDEN, A. (1986) *Intonation*. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics). Trad. cast. de I. Mascaró: *Entonación. Teoría general y aplicación al inglés*. Barcelona: Teide ( Serie Lingüística ), 1990

DECHERT, H. W. - RAUPACH, M. (Eds.) (1980) *Temporal Variables in Speech*,. Mouton: The Hague.

DI CRISTO, A. (1985) *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence: Université de Provence, Service des Publications.

EMERARD, F.- MORTAMET, L.- COZANNET, A. (1992) "Prosodic processing in a text-to-speech synthesis system using a database and learnig procedures", en G. BAILLY - C. BENOÏT (Eds.) *Talking Machines: Theories, Models and Designs*,. Amsterdam: Elsevier Science Publishers. pp. 225-254.

ENRÍQUEZ, E.- CASADO, C.- SANTOS, A. (1989) "La percepción del acento en español", *Lingüística Española Actual* 11: 241-269.

ESLING, J. H. (1988) "Computer coding of IPA symbols and detailed phonetic representation of computer data bases", *Journal of the International Phonetic Association* 18, 2: 99-106.

ESLING, J. H. - GAYLORD, H. (1993) "Computer codes for phonetic symbols", *Journal of the International Phonetic Association* 23, 2: 83-97.

ESTRUCH, M.- GARRIDO, J.M. (1995) "Análisis y clasificación de los contornos melódicos finales en un corpus de frases aisladas del español". Comunicación presentada en el *XXV Simposio de la Sociedad Española de Lingüística, Zaragoza, 11-14 de diciembre de 1995*. Resumen publicado en: *Revista Española de Lingüística* 26,1: 138-139.

FALLSIDE, F. (1985) "Frequency domain analysis of speech", en FALLSIDE, F.- WOODS, W.A. (Eds.) *Computer Speech Processing*. Englewood Cliffs, N.J.: Prentice Hall International. pp. 418-80

FANT, G. (1960) *Acoustic Theory of Speech Production*. Mouton: The Hague.

LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio. pp. 449-479.

- FANT, G. (1991). "Units of temporal organization. Stress groups versus syllables and words", en *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications. pp. 247-250.
- FARMER, A. (1984) "Spectrography", en CODE, C.- BALL, M. (Eds.) *Clinical Phonetics. Investigatory Techniques in Speech Pathology and Therapeutics*. London: Croom Helm. pp. 21-40
- FRENKENBERGER, S. *et al.* (1994) "Prosodic parsing based on parsing of minimal syntactic structures", en *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA*. pp. 143-146.
- FRY, D.B. (1979) *The Physics of Speech*. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics ).
- FUJISAKI, H. (1991) "Modelling the generation process of F0 contours as manifestation of linguistic and paralinguistic information", en *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications.
- FUJISAKI, H. - OHNO, S. - NAKAMURA, K. - GUIRAO, M. - GURLEKIAN, J. (1994) "Analysis of accent and intonation in Spanish based on a quantitative model", *Proceedings of the 1994 International Conference on Spoken Language Processing*. Vol. 1. pp. 355-358.
- GARRIDO, J.M. (1991) *Modelización de patrones melódicos del español para la síntesis y el reconocimiento de habla*. Bellaterra: Universitat Autònoma de Barcelona.
- GARRIDO, J.M. (1996) *Modelling Spanish Intonation for Text-to-Speech Applications*. Tesis doctoral. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona
- GARRIDO, J.M.- LLISTERRI, J.- de la MOTA, C.- RÍOS, A. (1993) "Prosodic differences in reading style: Isolated vs. Contextualized Sentences", en *Eurospeech'93. 3rd European Conference on Speech Communication and Technology. Berlin, Germany, 21-23 September 1993*. Vol 1. pp. 573-576.
- GARRIDO, J.M.- LLISTERRI, J.- MARÍN, R.- de la MOTA, C.- RÍOS, A. (1995) "Prosodic markers at syntactic boundaries in Spanish", en ELENIUS, K.- BRANDERUD, P. (Eds.) *ICPhS 95, Proceedings of the XIIIth International Congress of Phonetic Sciences. Stockholm, Sweden, 13-19 August, 1995*. Vol. 2. pp. 370-373.
- GILI FIVELA, B.- QUAZZA, S. (1996) "A Prosodic Parser for an Italian Text-to-Speech System", en *Actas del XII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, Sevilla, septiembre de 1996. Procesamiento del Lenguaje Natural* 19: 189-200.
- GOLD, B.- RABINER, L. (1969) "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain", *Journal of the Acoustical Society of America* 46: 442-448.
- GOLDMAN-EISLER, F. (1972) "Pauses, Clauses, Sentences", *Language and Speech* 15: 103-113.
- GROSS, M. (1975) *Méthodes en syntaxe. Régime des constructions complétives*. Paris: Hermann.
- HARMEGNIES, B.- POCH-OLIVÉ, D. (1992) "A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish", *Speech Communication* 11, 4-5: 429-438.
- HART, J. t'- COLLIER, R.- COHEN, A. (1990) *A Perceptual Study of Intonation. An Experimental - Phonetic Approach to Intonation*. Cambridge: Cambridge University Press. (Cambridge Studies in Speech Science and Communication )
- HESS, W. (1983) *Pitch Determination of Speech Signals*. New York: Springer Verlag.
- HIRSCHBERG, J.- PRIETO, P. (1994) "Training intonational phrasing rules automatically for English and Spanish text-to-speech", en *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA*. pp. 159-162
- HIRST, D.J. - ESPESSER, R. (1993) "Automatic modelling of fundamental frequency using a quadratic spline function", *Travaux de l'Institut de Phonetique d'Aix* 15: 71-85.
- HOLMES, J.N. (1988) *Speech Synthesis and Recognition*. Wokingham: Van Nostrand Reinhold (Aspects of Information Technology ).
- LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informàtica. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informàtica, Departamento de Filología Española, Universidad Autònoma de Barcelona - Editorial Milenio. pp. 449-479.

- HOWARD, H. - GOLDMAN, R.P. (1994) "From Text to Syllable in Castilian", en *Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, Córdoba, 20-22 julio 1994, Universidad de Córdoba*.
- HUALDE, J. I. (1991) "On Spanish Syllabification", en H. CAMPOS - F. MARTÍNEZ GIL (Eds.) *Current Studies in Spanish Linguistics*. Washington, D.C.: Georgetown University Press. pp. 475-494.
- IGLESIAS, J.L. (1994). *La duración de consonantes oclusivas y de consonantes aproximantes*. Manuscrito no publicado. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.
- JAVKIN, H.R. (1996) "Speech analysis and synthesis", en LASS, N.J (Ed.) *Principles of Experimental Phonetics*. St Louis: Mosby. pp. 245-276.
- JIMÉNEZ (1994) *Implementació d'un mètode d'estilització de corbes melòdiques*. Manuscrito no publicado. Barcelona: Enginyeria La Salle, Universitat Ramon Llull.
- KELLER, E. (Ed.) (1994) *Fundamentals of Speech Synthesis and Speech Recognition. Basic Concepts, State of the Art and Future Challenges*. Chichester: John Wiley & Sons.
- KENT, R.D.- READ, Ch. (1992) *The Acoustic Analysis of Speech*. London - San Diego: Whurr Publishers - Singular Publishing Group.
- KLATT, D.H. (1987) "Review of Tex-to-Speech Conversion for English", *Journal of the Acoustical Society of America* 82,3: 737-793; en ATAL, B.S.- MILLER, L.J.- KENT, R.D. (Eds.) (1991) *Papers in Speech Communication: Speech Processing*. New York: Acoustical Society of America. pp. 57-114.
- LADD, D. R. (1988) " 'Declination reset' and the hierarchical organization of units", *Journal of the Acoustical Society of America* 84, 2: 530-544.
- LADD, D.R. (1987) "A Model of Intonational Phonology for Use in Speech Synthesis by Rule", en LAVER, J.- JACK, M.A. (Eds.) *European Conference on Speech Technology. Edinburgh, September 1987*. Edinburgh: CEP Consultants Ltd. pp. 21-24
- LADEFOGED, P. (1996) *Elements of Acoustic Phonetics*. Chicago - London: University of Chicago Press. Second Edition.
- LAPORTE, E. (1988) *Méthodes algorithmiques et lexicales de phonétisation de textes*. Tesis doctoral. Centre d'études et de recherches en informatique linguistique, Université de Paris 7.
- LEHISTE, I. (1979) "Perception of Sentence and Paragraph Boundaries", en LINDBLOM, B. - ÖHMAN, S. (Eds.) *Frontiers of Speech Communication Research*. London: Academic Press. pp. 191-201.
- LEVELT, W.J.M. (1989) *Speaking. From Intention to Articulation*. Cambridge, Mass.: The MIT Press (ACL-MIT Press Series in Natural Language Processing)
- LINDBLOM, B. (1986) "On the origin and purpose of discreteness and invariance in sound patterns", en J.S PERKELL- D.H. KLATT (Eds.) *Invariance and Variability in Speech Processes*. Hillsdale: Lawrence Erlbaum Ass. pp. 493-523.
- LINDBLOM, B. (1990) "Explaining Phonetic Variation: A Sketch of the H and H Theory", en HARDCASTLE, W.J.- MARCHAL, A. (Eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers (NATO ASI Series D: Behavioural and Social Sciences, vol 55) pp. 403-439.
- LLISTERRI, J. (1991) *Introducción a la fonética: el método experimental*. Barcelona: Anthropos (Autores, Textos y Temas, Lingüística, 3).
- LLISTERRI, J.- MARÍN, R.- MOTA, C. de la - RÍOS, A. (1995) "Factors affecting F0 peak displacement in Spanish", en *Eurospeech'95. 4th European Conference on Speech Communication and Technology. Madrid, Spain, 18-21 September, 1995*. Vol 3. pp. 2061-2064.
- LÓPEZ, E. (1993) *Estudio de técnicas de procesamiento lingüístico y acústico para sistemas de conversión texto-voz en español basados en concatenación de unidades*. Tesis doctoral. Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid.
- LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autònoma de Barcelona - Editorial Milenio. pp. 449-479.

LÓPEZ, E. - ÁLVAREZ, J. - HERNÁNDEZ, L.A (1994) "Metodología para el modelado prosódico de un sistema de conversión de texto a habla en castellano", en *Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural. Córdoba, 20-22 julio 1994, Universidad de Córdoba*.

MACARRÓN, A. - ESCALADA, G.- RODRÍGUEZ, M.A. (1991) "Generation of duration rules for a Spanish text-to-speech synthesizer", en *Eurospeech'91. 2nd European Conference on Speech Communication and Technology. Genova, Italy, 24-26 September 1991*. pp. 617-620.

MACHUCA, M. (1991) *Estudio de las consonantes nasales del español en habla espontánea y en habla de laboratorio*. Trabajo de investigación de tercer ciclo. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

MARÍN, R. (1994) "Diseño y evaluación de un modelo de duración vocálica del español para la síntesis del habla", en *Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, Córdoba, 20-22 de julio de 1994, Universidad de Córdoba*.

MARÍN, R. (1995) "La duración vocálica en español", *Estudios de Lingüística* (Alicante), 10: 213-226.

MARTÍ, J. (1988) "FFT como herramienta de análisis en fonética", *Estudios de fonética experimental* 3: 233-251

MARTÍ, J. - NIÑEROLA, D. (1987) "SINCAS: un conversor texto-voz en castellano", en *Actas del III Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, Procesamiento del Lenguaje Natural, Boletín nº 5*: 112-122.

MARTÍNEZ CELDRÁN, E. (1985) "Cantidad e intensidad de los sonidos obstruyentes del castellano: hacia una caracterización acústica de los sonidos aproximantes" *Estudios de fonética experimental I*. Barcelona: PPU.

MARTÍNEZ CELDRÁN, E. (1993) "Nuevos datos sobre la dentalización de /s/", comunicación presentada en el *XXIII Simposio de la Sociedad Española de Lingüística*, Universidad de Lleida, 13-16 diciembre. Resumen publicado en: *Revista Española de Lingüística* 26,1.

MONAGHAN, A. I. C. (1992) "Heuristic strategies for higher level analysis of unrestricted text", en G. BAILLY- C. BENOÎT (Eds.) *Talking Machines: Theories, Models and Designs*. Amsterdam: Elsevier Science Publishers. pp. 143-162.

MOTA, C. de la - RÍOS, A. (1995) "Problemas en torno a la transcripción fonética del español: los alfabetos fonéticos propuestos por IPA y RFE y su aplicación a un sistema automático", *Acta Universitatis Wratislaviensis* nº 1660, Estudios Hispánicos IV. Wroclaw. pp. 97-109.

MOTA, C. de la (1991) "A study of [r] and [ʀ] in spontaneous speech", en *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications. Vol 4. pp. 386-389.

MOTA, C. de la (1995) *La representación gramatical de la información nueva en el discurso*. Tesis doctoral. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

NAVARRO TOMÁS, T. (1916) "Cantidad de las vocales acentuadas", *Revista de Filología Española* 3: 387-407.

NAVARRO TOMÁS, T. (1918) "Diferencias de duración entre las consonantes españolas", *Revista de Filología Española* 5: 367-393

NAVARRO TOMÁS, T. (1918) *Manual de pronunciación española*. Madrid: CSIC. 21ª edición, 1982.

NAVARRO, T. (1944) *Manual de entonación española*.. Madrid: Guadarrama. 4ª edición.

NESPOR, M.- VOGEL, I. (1983). "Prosodic Structure above the Word", en A. CUTLER - R. D. LADD (Eds.) *Prosody. Models and Measurements*. Heidelberg: Springer Verlag. pp. 123-140.

NOOTEBOOM, S.G. (1991). "Some observations on the temporal organisation and rhythm of speech", en *Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France*. Aix-en-Provence: Université de Provence, Service des Publications. pp. 228-237.

LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio. pp. 449-479.

NOOTEBOOM, S.G. *et al.* (1978) "Contributions of prosody to speech perception", en W. J. M. LEVELT - G. G. FLORES D'ARCAIS (Eds.) *Studies in the Perception of Language*. Chichester: John Wiley. pp. 75-107.

O'SHAUGHNESSY, D. (1987) *Speech Communication. Human and Machine*. Reading, Mass.: Addison Wesley.

O'SHAUGHNESSY, D. (1990). "Relations between syntax and prosody for speech synthesis", *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France.

PÉREZ, J.C.- VIDAL, E. (1991) "Un sistema de conversión de texto a voz para el castellano", *Sociedad Española para el Procesamiento del Lenguaje Natural, Boletín* nº 11: 197-208.

PIERREHUMBERT, P. (1980). *The Phonology and Phonetics of English Intonation*. Tesis doctoral. Massachusetts Institute of Technology. Bloomington: Indiana University Linguistics Club, 1987.

QUENÉ, H. - KAGER, R. (1992). "The derivation of prosody for text-to-speech from prosodic sentence structure", *Computer Speech and Language* 6: 77-98.

QUILIS, A. (1966) "Sobre los alófonos dentales de /s/, *Revista de Filología Española*, XLIX: 335-343.

QUILIS, A. (1981) *Fonética acústica de la lengua española*. Madrid: Gredos (Biblioteca Románica Hispánica, Manuales, 49)

QUILIS, A. (1985) *El comentario fonológico y fonético de textos. Teoría y práctica*. Madrid: Arco/Libros.

QUILIS, A. (1993) *Tratado de fonología y fonética españolas*. Madrid: Gredos (Biblioteca Románica Hispánica, Manuales, 74)

QUILIS, A.- FERNÁNDEZ, J.A. (1964) *Curso de fonética y fonología españolas para estudiantes angloamericanos*. Madrid: Consejo Superior de Investigaciones Científicas (Collectanea Phonetica 2). 10ª edición, 1982.

RILEY, M.D. (1992) " Tree-based modelling of segmental durations", en BAILLY, G.- BENOÎT, C. (Eds) *Talking Machines. Theories, Models and Designs*. Amsterdam: North-Holland / Elsevier Science Publishers. pp. 265-274.

RÍOS, A. (1991) *Caracterización acústica del ritmo del castellano*. Trabajo de investigación de tercer ciclo. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

RÍOS, A. (1993) "La información lingüística en la transcripción fonética automática del español", *Sociedad Española para el Procesamiento del Lenguaje Natural*., *Boletín* nº 1 : 381-387.

RODRÍGUEZ, M.A.- ESCALADA, J.G.- MACARRÓN, A.- MONZÓN, L. (1993) " AMIGO: Un conversor texto-voz para el español", *Sociedad Española para el Procesamiento del Lenguaje Natural*, *Boletín* nº 13: 389-400.

SAGER, J.C. (1992) "La industria de la lengua", in SAGER, J.C. *La industria de la lengua, La lingüística computacional - los trabajos de UMIST. La traducción especializada y su técnica*. Barcelona: Servei de Llengua Catalana, Universitat de Barcelona. pp. 7-29.

SANTEN, J.P.H. van (1992) "Deriving text-to-speech durations from natural speech", en BAILLY, G.- BENOÎT, C. (Eds.) *Talking Machines. Theories, Models and Designs*. Amsterdam: North-Holland / Elsevier Science Publishers. pp.265-274

SANTEN, J.P.H. van - OLIVE, J.P. (1990). "The analysis of contextual effects on segmental duration", *Computer Speech and Language* 4: 359-361.

SANTOS, A.- MUÑOZ, P.- MARTÍNEZ, M. (1988) "Diseño y evaluación de reglas de duración en la conversión de texto a voz", *Procesamiento del Lenguaje Natural, Boletín* nº 6: 69-92.

SAWUSCH, J.R. (1996) "Instrumentation and methodology for the study of speech perception", en LASS, N.J (Ed.) *Principles of Experimental Phonetics*. St Louis: Mosby. pp. 525-550.

SCHARPFF, P. J. - HEUVEN, V.J. van (1988). "Effects of pause insertion on the intelligibility of low quality speech", en *Proceedings of the 7th FASE symposium (Speech'88)*, *Edinburgh*. Vol. 1. pp. 261-268.

LLISTERRI, J.- AGUILAR, L.- GARRIDO, J.M.- MACHUCA, M.J.- MARÍN, R.- DE LA MOTA, C.- RÍOS, A. (1999) "Fonética y tecnologías del habla", in BLECUA, J.M.- CLAVERÍA, G.- SÁNCHEZ, C.- TORRUELLA, J. (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio. pp. 449-479.

- STEVENS, K.N. (1989) "On the quantal nature of speech", *Journal of Phonetics* 17, 1/2: 3-45; en KENT, R.D.- ATAL, B.S.- MILLER, J.L. (Eds.) (1991) *Papers in Speech Communication: Speech Production*. New York: Acoustical Society of America. pp. 357-399
- STONE, M. (1996) "Instrumentation for the study of speech physiology", en LASS, N.J (Ed) *Principles of Experimental Phonetics*. St Louis: Mosby. pp. 495-524
- TAKEFUTA, Y. (1975) "Method of Acoustic Analysis of Intonation", en SINGH, S. (Ed) *Measurement Procedures in Speech, Hearing and Language*. Baltimore: University Park Press. pp. 363-378.
- VIDAL BENEYTO, J. (Dir) (1991) *Las industrias de la lengua*. Trad. de M. Alvar Ezquerro *et al.* Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide (Biblioteca del Libro, 5).
- WAKITA, H.J. (1996) "Instrumentation for the study of speech acoustics", en LASS, N.J (Ed.) *Principles of Experimental Phonetics*. St Louis: Mosby. pp. 469-494.
- WANG, M. Q. - HIRSCHBERG, J. (1992). "Automatic classification of intonational phrase boundaries", *Computer Speech and Language* 6: 175-196.
- WELLS, J. (1987) "Computer-coded phonetic transcription", *Journal of the International Phonetic Association* 17, 2: 94-114.
- WELLS, J. (1990) "Computer-coded Phonemic Notation of Individual Languages of the European Community", *Journal of the International Phonetic Association* 19, 1: 31-54.
- WELLS, J.C. (1995) "Computer-coding the IPA: a proposed extension of SAMPA". Publicación electrónica. URL: <http://www.phon.ucl.ac.uk/home.sampa/home/x-sampa.html>