



Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnología de la Información



TRABAJO DE DIPLOMA

“Implementación de un segmentador automático de habla fluida”

Autor: Adilson Pérez González

Tutor: Dra. Maria E. Hernández – Díaz Huici

Santa Clara

2008

"Año 50 de la Revolución"



Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnología de la Información



TRABAJO DE DIPLOMA

“Implementación de un segmentador automático de habla fluida”

Autor: Adilson Pérez González

Tutor: Dra. Maria E. Hernández – Díaz Huici

Profesor titular

Centro de Estudios de Electrónica y Tecnología de la Información

Facultad de Ingeniería Eléctrica

E-mail: mariae@uclv.edu.cu

Santa Clara

2008

"Año 50 de la Revolución"



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería en Telecomunicaciones y Electrónica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Autor

Firma del Jefe de Departamento
donde se defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

PENSAMIENTO

El pensamiento es más que un derecho, es el aliento mismo del hombre.

Victor Hugo

DEDICATORIA

A todos y cada uno de mis familiares, a mi esposa que con su desvelo y paciencia aportó un grano de arena en la confección de este trabajo, a mi mamá que contribuyó con sus conocimientos y experiencias, a mi tío y prima que me facilitaron algunos materiales útiles para el mismo y a mi papá que aportó la ayuda necesaria.

AGRADECIMIENTOS

Deseo expresar mi agradecimiento a los profesores de la Universidad Central, sin cuyo esfuerzo y dedicación no hubiera sido posible la adquisición de conocimientos durante el transcurso de la carrera. Especialmente agradezco la atención brindada por la tutora que siempre estuvo en la mejor disposición de facilitarme su apoyo y sin la cual este trabajo no se habría podido llevar a cabo.

TAREA TÉCNICA

- Revisión bibliográfica para conocer las diferentes características presentes en el habla fluida y los tipos de segmentadores utilizados en el análisis de la misma.
- Adquisición de habilidades con el MATLAB y con los diferentes software libres asociados al tema, para obtener las variaciones acústicas necesarias y encontrar las características distintivas.
- Programación de un segmentador para situar marcas en la señal del habla en lugares donde ocurren cambios espectrales de interés.
- Análisis del desempeño del algoritmo programado teniendo en cuenta oraciones en idioma Inglés y en Español.
- Confección del informe final.

Firma del Autor

Firma del Tutor

RESUMEN

El estudio y análisis del habla tienen una gran importancia para el ser humano por ser esta su principal forma de comunicación. Es por ello que en la actualidad se buscan algoritmos y técnicas que faciliten su desempeño.

El presente trabajo parte de un análisis de los distintos aspectos de la prosodia y de los diferentes tipos de modificaciones que se producen en el habla fluida que dan lugar a los cambios espectrales.

Dado estos planteamientos, se lleva a cabo un análisis de varios tipos de segmentadores, validados en la literatura, en cuanto a sus características, para realizar la implementación de un segmentador a partir de un algoritmo seleccionado, que permita colocar en la señal de audio marcas en los puntos donde ocurren cambios espectrales de interés.

Como aspecto novedoso, se efectúa de forma automática las localizaciones de los puntos en la señal del habla donde ocurren variaciones importantes.

Por último, se ilustran las pruebas realizadas al algoritmo implementado mediante una serie de oraciones tanto en idioma Inglés como en Español donde se destacan los cambios espectrales que ocurren durante el habla fluida y su consideración como puntos de inflexión a tener en cuenta en un análisis prosódico posterior.

TABLA DE CONTENIDOS

PENSAMIENTO	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
TAREA TÉCNICA.....	iv
RESUMEN	v
INTRODUCCIÓN.....	1
CAPÍTULO 1. SELECCIÓN DE SEGMENTADOR PARA PROSODIA	4
1.1 Características acústicas del habla normal.....	4
1.2 Análisis suprasegmental del habla	4
1.3 Modificaciones en el habla fluida	7
1.3.1 Co-Articulación como base para la asimilación	8
1.3.2 Asimilación en el habla normal	9
1.3.3 Procesos fonológicos de asimilación	9
1.4 Tipos de segmentadores	9
CAPÍTULO 2. MATERIALES Y MÉTODOS.....	14
2.1 Materiales	14
2.2 Métodos.....	15
2.2.1 Procesamiento general del algoritmo seleccionado	15

2.2.2	Modificaciones realizadas e implementación	23
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN		26
3.1	Validez y funcionalidad de las marcas	26
3.2	Observaciones en el habla fluida.....	31
CONCLUSIONES Y RECOMENDACIONES		36
REFERENCIAS BIBLIOGRÁFICAS		38
ANEXOS		41
Anexo A	Funciones implementadas	41
Anexo B	Listado del programa en MATLAB versión 6.5	46

INTRODUCCIÓN

Tanto en estudios lingüísticos como en estudios clínicos son de primordial interés los análisis objetivos relacionados con la prosodia (Kent & Bale, 2000; Epstein, 2002). Sin embargo, debido a la complejidad que implica el análisis de unidades lingüísticas mayores que un fonema, aparecen muchos trabajos relacionados con la calidad vocal en la literatura y una cantidad muy reducida se reporta para el análisis de fenómenos articulatorios y de prosodia (Ferrer, 2005).

Las técnicas de Procesamiento de Voz han sido empleadas en ambos campos para documentar principalmente los fenómenos relacionados con la calidad vocal, como por ejemplo, mediciones relacionadas con el período fundamental, mediciones de ruido y en la confección de espectrogramas además de otras formas de visualización. Para ello se han hecho suposiciones fundamentales, tales como, considerar la señal estacionaria para intervalos cortos de tiempo (de 5 a 20ms). Esta suposición introduce cierta incertidumbre ya que tanto el período como la amplitud de la señal pueden variar dentro de ese intervalo. Cuando el segmento bajo análisis se hace mayor (de 20 a 100ms) la incertidumbre aumenta, además, por los cambios en calidad del sonido y las transiciones entre sonidos. En segmentos mayores (oraciones) el grado de complejidad se incrementa mucho más y algunas de las suposiciones necesarias no son ya válidas (Xin, 2008). Es por ello que en la actualidad se buscan algoritmos y técnicas que faciliten el análisis de segmentos con suficiente duración para estudiar la prosodia y en general otras características en el habla fluida, que representan la forma de comunicación natural.

La tarea planteada se enmarca en las investigaciones que lleva a cabo el Laboratorio de Procesamiento de Voz del CEETI con la finalidad de encontrar medidas que apoyen el diagnóstico y evaluación de los pacientes con dificultades en la comunicación oral.

El resultado de este trabajo está encaminado a facilitar la aplicación de las herramientas de análisis, desarrolladas por este colectivo, en el habla fluida.

Una vez que la población cubana va camino al envejecimiento y la atención primaria está bien establecida, los retos se dirigen a lograr mejor calidad de vida para las personas con algunas limitaciones en su comunicación que impiden un exitoso desempeño como seres sociales.

Lo anterior conduce a la siguiente formulación del problema:

Una forma de procesar el habla fluida, es segmentar la oración teniendo en cuenta aquellos sitios donde ocurren cambios espectrales de interés. Técnicas similares se han empleado en sistemas de reconocimiento y verificación de locutor (Liu, 1996; Salomon, 2004; Susand & Rabiner, 2006).

En la literatura aparecen diferentes tipos de segmentadores basados en distintos parámetros. Para facilitar el análisis del habla, se necesita un sistema de segmentación que sitúe marcas como guía para localizar las características distintivas de la señal acústica asociadas a los cambios suprasegmentales.

Para resolver este problema se debe elegir un algoritmo apropiado e implementarlo para su posterior uso.

De este problema se originan las siguientes interrogantes científicas:

- ¿Cómo seleccionar un algoritmo para la segmentación del habla basado en características distintivas?
- ¿Cómo evaluar la efectividad de la implementación realizada?

Por lo tanto el objetivo general del presente proyecto es:

Implementar un segmentador, basado en un algoritmo validado en la literatura, que permita situar en la señal de audio (archivo.wav) marcas temporales en un segmento del habla fluida en los puntos donde ocurren cambios espectrales de interés.

De este objetivo general se derivan los objetivos específicos siguientes:

- Seleccionar el algoritmo adecuado.

- Programar el algoritmo seleccionado (o parte de él) en MATLAB versión 6.5.
- Evaluar los resultados alcanzados.

Con este proyecto se contribuye al desarrollo de procedimientos para un mejor análisis e identificación de variaciones en el habla fluida.

La implementación de estas soluciones permitirá a los especialistas e investigadores realizar estudios y análisis comparativos, que contribuirán a una mejor comprensión del mecanismo de producción y percepción del habla.

El informe de la investigación se estructura en introducción, capitulario, conclusiones, recomendaciones, referencias bibliográficas y anexos.

En la introducción se deja definida la importancia, actualidad y necesidad del tema que se aborda.

El Capítulo 1 se dedica a la caracterización del problema que se aborda a partir de un análisis de la literatura. Se comentan varios tipos de segmentadores para el análisis del habla y se realiza una selección.

El Capítulo 2 se utiliza para expresar el diseño metodológico de la investigación y el algoritmo seleccionado.

En el Capítulo 3 se realiza la validación de la efectividad del método mediante la comparación de los resultados obtenidos con ejemplos prácticos.

CAPÍTULO 1. SELECCIÓN DE SEGMENTADOR PARA PROSODIA

En este capítulo se hace un análisis de los diferentes aspectos de la prosodia, así como de las diferentes características acústicas del habla normal y los tipos de modificaciones que le ocurren a las articulaciones en el habla fluida, debido a la gran importancia que juegan estos factores en los distintos cambios espectrales que ocurren en la señal del habla. Además se realiza un estudio de los diferentes tipos de segmentadores validados en la literatura para una posterior selección del algoritmo a implementar.

1.1 Características acústicas del habla normal

El habla, como medio de comunicación humana, porta información en varios niveles: lingüístico, para-lingüístico y extra-lingüístico (Laver, 1991). En el primero se incluye la información de lo que se desea transmitir, el “texto” del mensaje. El segundo porta información sobre los estados de ánimo, emocionales u otros de la persona, así como su procedencia social y otros factores culturalmente determinados. El tercer nivel comprende todos los factores físicos y fisiológicos (incluyendo todos los aspectos orgánicos involucrados en la producción del habla) característicos del locutor.

Dada la cantidad de información que porta el habla y la importancia que reviste para los seres humanos, por ser la vía fundamental de comunicación, se comprende la amplitud de las investigaciones que se dedican al estudio del habla en estos tres niveles.

1.2 Análisis suprasegmental del habla

La entonación, la razón del habla, la unión, la acentuación y el ritmo son aspectos de la producción del habla que trascienden los límites de fonemas individuales, ellos son clasificados colectivamente como *Prosodia* y requieren un análisis suprasegmental. Estas

características suprasegmentales no pueden estar descritas con referencia a un segmento sencillo del sonido, pero pueden causar efectos en al menos una sílaba o en segmentos de diferentes partes de una expresión. Algunas personas que estudian la fonética usan el término *Prosodia* para encerrar estos aspectos de análisis fonéticos y fonológicos (Cristal, 1982). No importa qué término es usado, una comprensión completa de las características del habla normal y anormal requiere de un análisis de entonación, razón del habla, unión, acentuación y ritmo.

Entonación (Intonation)

La entonación es la variación sistemática en el tono de la sonoridad que produce diferencias significativas en las unidades del habla como pueden ser frases y oraciones. El tono es la percepción de la frecuencia. Para comprender la entonación, el tono es la percepción de la frecuencia fundamental, la frecuencia de vibración de las cuerdas vocales. En (Cristal, 1982) se han identificado cinco funciones para la entonación en Inglés: gramatical, semántica, de emoción, social y psicológica.

Las variaciones en la entonación involucran cambios en el tono y eso incorpora cambios en la duración y en la acentuación. La percepción de variaciones de tonos es relativa, pues se basa en variaciones relativas de F_0 . De otra forma, los cambios significativos de tonos no se podrían interpretar de la misma manera cuando hablan hombres, mujeres y niños, cuya frecuencia fundamental se encuentra entre 120 y 130 Hz, entre 190 y 255 Hz y sobre los 265 Hz respectivamente (Tiffany & Carrell, 1977; Cruttenden, 1986). Si, por ejemplo, un tono fijo en vez de uno alto fuera el indicativo de una pregunta, el sistema de entonación no sería funcional. La frecuencia fundamental incrementa cuando se realiza una pregunta, decrece para una afirmación completa y es constante para afirmaciones incompletas. La frecuencia fundamental decrece en el siguiente orden: niños, mujeres y hombres.

Razón del habla (Tempo)

Esta puede ser definida como la percepción del índice o la velocidad de producción del habla. La razón del habla de las personas puede ser medida en sílabas por segundo o en palabras por minuto, pero la sílaba es una mejor medida ya que mantiene una razón más estándar en un tramo. No importa la medida que sea usada, las diferencias en el habla de varias personas reflejan cambios en la duración de la producción de sonidos y las pausas

entre ellos, ambos resumen cómo la velocidad del habla incrementa y disminuye. Es por ello que la razón del habla puede ser identificada como: índice o velocidad del hablante, duración de sonidos y sílabas, duración de pausas y localización (Ohde & Sharf, 1992).

Unión (Juncture)

Una de las principales consecuencias de los cambios en la razón del habla son los cambios en la duración de pausas y el alargamiento y estrechamiento con razón disminuida e incrementada respectivamente (Ohde & Sharf, 1992). Dos transiciones diferentes del habla son: la transición entre sílabas de la misma palabra o entre palabras vecinas (unión interna) que normalmente no involucra una pausa y las diferenciaciones pueden estar hechas entre pares mínimos y la transición entre sílabas que marcan los límites entre los contornos de entonación (unión externa).

Acentuación (Stress)

Es la prominencia relativa o un énfasis en la sílaba de una palabra o en unidades largas del habla como frases y oraciones. Aunque existe un convenio general sobre el patrón de la acentuación sobre las palabras, hay desacuerdos acerca de la base para esta diferenciación. Se pueden esperar sílabas acentuadas para tener una alta frecuencia fundamental, una duración más larga y una intensidad mayor que las sílabas sin acentuar, pero esta es una excepción dentro de las generalidades (Lehiste, 1970; Ladefoged, 1975). En frases y oraciones, la localización de la acentuación máxima depende del significado propuesto.

Ritmo (Rythm)

El ritmo es el patrón de la localización de la acentuación y de la duración de los intervalos entre sílabas acentuadas (Ohde & Sharf, 1992). Los patrones usuales para el idioma Inglés son: sílabas alternadas periódicamente acentuadas con la presencia de una o más sílabas sin acento. Estudiosos de fonética describen el idioma Inglés como un lenguaje de una acentuación regulada. Otros idiomas como el Francés son descritos como de sílabas reguladas porque cada sílaba es aproximadamente de igual duración y las vocales no son generalmente acentuadas. La lengua castellana se caracteriza por construir unidades rítmicas de cinco a diez sílabas; y de ellas las más frecuentes cuentan con siete u ocho (Gili Gaya, 2001).

La implicación del término acentuación regulada es que el ritmo del habla involucra el control de la localización de acentos y los intervalos de tiempo entre ellos. En el Inglés existe un patrón de ritmo sincrónico, es un lenguaje que presenta generalmente iguales intervalos entre sílabas acentuadas.

1.3 Modificaciones en el habla fluida

El concepto de distinción o diferencia presupone el de contraste o el de oposición. Una cualidad fónica tendrá función distintiva cuando se oponga a otra, o sea cuando ambos formen una oposición fonética. Las diferencias fónicas que en una lengua dada permiten distinguir las significaciones son oposiciones fonológicas distintivas o relevantes. A las unidades fonológicas indivisibles se les llama fonemas, los cuales se clasifican en vocales y consonantes.

Las características distintivas describen los sonidos del habla en un nivel subsegmental, tienen una relación relativamente directa para la acústica y la articulación. Toman valores binarios y forman un set mínimo que puede distinguir cada segmento de todos los otros (Jakobson, Fant & Halle, 1952). Además pueden describir muchas de las variaciones contextuales de un segmento las cuales son: estilos de hablas individuales y asimilación fonológica a través de los confines de palabras.

En el habla fluida se ponen de manifiesto con mayor claridad estas funciones distintivas, apreciando con ello la influencia, es decir, las diferentes modificaciones que ejercen unos fonemas sobre otros, de ahí su importancia para el análisis de voz tanto en el habla normal como patológica. La naturaleza de la operación de los sistemas articulatorios es la responsable de las modificaciones contextuales.

En el habla fluida aparecen las siguientes modificaciones a las articulaciones:

- Adaptación: Modificación que sufre un fonema como resultado de la influencia de los sonidos vecinos, los cuales cambian ligeramente la forma del tracto vocal y las posiciones articulatorias.
- Asimilación: Son los casos extremos de adaptación.
- Co-Articulación: Dos articuladores se mueven al mismo tiempo preparándose para la producción de dos fonemas diferentes.

Como resultado de la combinación de adaptación y co-articulación el habla resultante es más robusta porque la información de una simple articulación perdura más.

1.3.1 Co-Articulación como base para la asimilación

La co-articulación presenta una gran importancia para los estudios fonéticos. El habla involucra un complejo sistema articulatorio con la interacción de un gran número de componentes que son los responsables para la producción de ciertas propiedades. No está claro cómo son en realidad estas propiedades específicas ni cómo interactúan los componentes involucrados en la producción de las mismas, pero han existido diferentes teorías acerca de la configuración del sistema (Ohde & Sharf, 1992).

La co-articulación incluye cambios anticipados durante la producción del habla, limitaciones en la sensibilidad de estructuras articulatorias en cambios de posición entre sonidos. Cualquier mecanismo durante la co-articulación es probablemente la base de cambios en la pronunciación de palabras y de expresiones que son reconocidas como ejemplo de asimilación.

No todos los estudiosos de la fonética están de acuerdo con la relación entre co-articulación y asimilación. El proceso de cambios de sonidos y de cambios que ocurren debido a ellos, han estado definidos como co-articulación (Shriberg & Kent, 1982) y asimilación (Bronstein, 1960) respectivamente. La co-articulación puede ser vista como la base para la asimilación (Tiffany & Carrell, 1977). Por otra parte la asimilación y la co-articulación pueden ser consideradas como dos formas diferentes de un proceso de acomodamiento en los cambios de sonidos (MacKay, 1987). La falta de una clara comprensión sobre los procesos responsables para los cambios de sonidos está vinculada con las diferencias, sin resolver, relacionadas con la naturaleza del control motor del lenguaje. Algunos investigadores asumen que la articulación es regulada del mismo modo que los movimientos de las extremidades (Kelso, Saltzman, & Tuller, 1986). Otros plantean que la articulación es determinada por algunas extensiones de representaciones fonéticas en el sistema nervioso central, el cual puede ser influenciado por cada lenguaje en específico (Smith & Mclean-Muse, 1986, 1987; Flege, 1988).

1.3.2 Asimilación en el habla normal

En el habla normal la asimilación puede implicar características del lugar, de sonoridad y de la manera de producción (Ohde & Sharf, 1992). Algunos de los cambios son bastante amplios en Inglés; los otros están limitados para una o dos palabras. Una cantidad de cambios están tan presentes en el habla normal que la mayoría de las personas no perciben la diferencia entre pronunciaciones de fonemas. Debería ser puntualizado que las asimilaciones ocurren más frecuentemente de forma casual que en el habla formal.

1.3.3 Procesos fonológicos de asimilación

Entre los varios procesos que ocurren durante el habla normal y anormal se encuentran aquellos que pueden estar explicados en términos de asimilación (Ingram, 1976).

Tanto para niños como adultos los cambios regresivos parecen ser más comunes que los progresivos. En el habla de los adultos todas las asimilaciones son adyacentes, mientras que en los niños predomina grandemente el caso contrario. Además, es raro encontrar asimilaciones completas en los adultos, siendo esto más común en los niños. Aunque todas las asimilaciones se corresponden con transiciones simplificadas entre los diferentes estados articulatorios que caracterizan los diferentes sonidos, esto se manifiesta más claro en los niños que en los adultos.

1.4 Tipos de segmentadores

En la literatura aparecen algunos algoritmos para segmentar las señales del habla basados en distintos parámetros. Una de las áreas más interesantes en el reconocimiento del habla es el problema de reconocimiento de dígitos. En un sistema de reconocimiento realizado por (Rabiner & Sambur, 1975) se explican las dos partes fundamentales del mismo, las cuales están interrelacionadas. La función de la primera parte es segmentar una serie de dígitos en dígitos individuales, la segunda parte es reconocer los dígitos individuales basados en los resultados de la segmentación. La segmentación de dígitos está basada en un análisis de sonoridad de la serie de dígitos, así como la información sobre la posición y la amplitud mínima en el contorno de energía de la expresión. La estrategia de reconocimiento de dígitos es similar al algoritmo usado por Sambur y Rabiner para dígitos aislados, pero con

modificaciones importantes debido a imprecisiones, con lo cual los límites exactos de los dígitos pueden estar localizados.

Por otra parte (Hess, 1975) en un trabajo realizado para un sistema digital sincrónico de extracción de picos (PDFES de sus iniciales en Inglés), orientado al reconocimiento de fonemas del habla, expone la necesidad de un segmentador unido a extracción de tonos como un primer paso y análisis de los formantes como el objetivo final. Usando tanto los niveles de la señal del habla y la señal diferenciada, como los parámetros en el dominio del tiempo, el algoritmo de segmentación deriva un parámetro de la señal el cual describe la rapidez de los movimientos articulatorios. Luego, la señal es dividida en segmentos estacionarios y transitorios, un segmento estacionario es asociado a un fonema. El procedimiento para encontrar formantes, se basa en un subconjunto de períodos de tonos que son seleccionados por el algoritmo de segmentación y transformados en el dominio de la frecuencia.

En la señal del habla los elementos (fonemas, palabras, etc.), no son claramente separados, se asocian unos dentro de otros. Una segmentación realiza una subdivisión en secciones consecutivas discretas. El primer paso de segmentación en el PDFES se realiza antes de la determinación de tonos. Por detección de dos características binarias: *silencio – habla* y *sonoro – sordo*, se ejecuta una presegmentación a secuencias de sílabas. Para realizar pasos de segmentación para fonemas, el sistema debe tener en cuenta el tracto vocal y movimientos articulatorios. Para cada fonema existe una posición articulatoria del tracto vocal. En secciones con movimientos relativamente rápidos se encuentran los segmentos dinámicos y en secciones con movimientos lentos aparecen los segmentos estacionarios.

Otras de las tantas utilidades de los segmentadores queda validada por (Mermelstein, 1975) quien aborda una estrategia para el análisis acústico del habla, donde cada proceso del análisis es aplicado a un segmento del habla etiquetado apropiadamente y los resultados son una posible subsegmentación del segmento original. Los segmentos resultantes del análisis son etiquetados de acuerdo a los resultados del procedimiento. Las ventajas de esta estrategia es que no son considerados más segmentos que esos, realmente diferenciados por los pasos del análisis. La extracción de pistas acústicas pertinentes a una característica fonética puede ubicar a clases y sonidos separados en la base de otras pistas y esto sirve

para incrementar la fiabilidad del segmento etiquetado. El análisis secuencial se basa en una estructura para unidades de sílabas de la señal del habla.

Una unidad de sílaba es definida como un segmento de la señal del habla delimitada por un suavizamiento del tiempo y ponderación de frecuencias en el espectro de la señal. El foco de atención está en la caracterización de segmentos en unidades de sílabas y la relación entre estos.

El algoritmo de segmentación propuesto por (Liu, 1996) es utilizado en un sistema de reconocimiento basado en características distintivas. Utiliza parámetros de frecuencia para dividir la señal en bandas. En cada banda una forma de onda de energía es construida, la derivada de la energía es calculada y los picos en las derivadas son detectados. Estos picos representan tiempos de cambios espectrales abruptos en las bandas. Todo lo anterior se efectúa teniendo en cuenta las categorías fonéticas, lo que contiene de manera implícita los diferentes aspectos de la producción del habla como son: la entonación, la razón del habla, las uniones, la acentuación y el ritmo. El algoritmo se basa en que un discurso de entrada pasa a través de una etapa general del procesamiento de la señal, de quien son las salidas que alimentan una etapa de procesamiento de marcas especificadas por tipo. La salida del procesamiento tipo-específico es una serie de marcas especificadas en tiempo y tipo.

Es de particular interés el trabajo reportado por (Salomon, 2004). Este plantea la detección de marcas en el habla mediante el uso de información temporal para un sistema de reconocimiento de voz. Utiliza segmentación para extraer información temporal lingüísticamente relevante. La información temporal es particularmente importante en la percepción humana del habla cuando la señal está altamente degradada. Los parámetros seleccionados incluyen energía onset y offset (calculadas, usando un algoritmo adaptativo) y medidas de contenido periódico y aperiódico.

Las motivaciones del trabajo de Salomon provienen, sobre todo, de estudios de percepción del discurso. Un problema principal en el desarrollo de sistemas de reconocimiento del habla es la detección del discurso a partir del ruido (Viikki, 2001) u otra forma de reducción espectral. El propósito radica en construir un detector para eventos acústicos, independientemente de la información espectral detallada.

Adicionalmente, la incorporación de parámetros temporales debe mejorar el desempeño y aumentar la resistencia ante ruidos, conocida la información espectral. Sin embargo, el desempeño mejorado, no puede ocurrir si el ruido tiene características temporales como el habla.

Este trabajo se fundamenta en los resultados de (Espy Wilson, 1992) y (Bitar, 1997). El primero empleó, el grado de brusquedad en una medida de diferencia de energía para distinguir la diferencia entre /w, y, r/ y consonantes sonoras nasales /m, n, l/, y Bitar reportó el uso de las diferencias de energía para encontrar ráfagas de consonantes oclusivas y realizar segmentación.

Además, tiene similitud con el procesamiento utilizado por (Browne & Cooke, 1994) donde se usó información de onsets, offsets y tonos como pistas en un sistema para el análisis auditivo de la escena.

En este estudio, la detección de acontecimientos se basa en el procesamiento general de todas las bandas de frecuencia y la adaptatividad del procesamiento de la señal. El cambio de energía está combinado con información acerca de la periodicidad y aperiodicidad para detectar acontecimientos. El análisis es realizado tanto para la señal limpia como para el empobrecimiento espectral de la misma.

Después del análisis realizado a los distintos tipos de segmentadores se observan las siguientes características: el segmentador de Rabiner y Sambur es diseñado para números, cuyas unidades incluyen una o dos sílabas y no marcan cambios espectrales de interés en el habla fluida. Mientras que el método de Hess es muy robusto pero se enfoca en el análisis de fonemas en particular y por lo tanto se pueden perder pistas de valor prosódico. Mermelstein para el desarrollo de su segmentador parte de segmentos ya etiquetados lo cual no corresponde con las características de la aplicación que se persigue. El segmentador propuesto por Liu, permite situar marcas en la oración seleccionada, en aquellos instantes de tiempo donde ocurren cambios espectrales notables, siendo de gran utilidad para el análisis del habla fluida y para prosodia. El algoritmo de Salomon pudiera emplearse como un post-procesamiento a la implementación que se tiene como propósito, que es la segmentación del habla fluida donde ocurren cambios espectrales importantes.

Conclusiones parciales

Teniendo en cuenta tanto el objetivo del trabajo como los tipos de segmentadores validados en la literatura, el algoritmo de Liu agrupa una serie de parámetros y desarrolla pasos de gran utilidad para localizar cambios espectrales significativos en la señal del habla fluida teniendo en cuenta los distintos aspectos de producción del habla descritos en este capítulo.

CAPÍTULO 2. MATERIALES Y MÉTODOS

En el desarrollo de este capítulo se dejan definidos los materiales y métodos utilizados para la realización del trabajo. Se exponen una serie de especificidades a tener en cuenta en el proceso de grabación de las diferentes oraciones que son analizadas, así como las distintas herramientas utilizadas. Se realiza una descripción del algoritmo seleccionado con la cual se explican sus ventajas y características, además se llevan a cabo una serie de modificaciones del mismo y se describen las diferentes funciones implementadas para llegar a los resultados esperados.

2.1 Materiales

Con el propósito de localizar de forma automática cambios espectrales de interés en la señal del habla se tienen en cuenta parámetros de frecuencia y energía. Para el cálculo de los mismos se contó con diferentes oraciones tanto en idioma Inglés como en Español.

Estas fueron digitalizadas empleando la tarjeta de sonido (SoundBlaster, Creative Labs Inc) de la PC. La adquisición se efectuó a través del utilitario “Sound Recorder” del sistema operativo Windows. Se empleó una frecuencia de muestreo de 16000 Hz y una resolución de 8 bits con un solo canal (mono) a 15kb/seg almacenándose las grabaciones en ficheros de audio con extensión .WAV, en formato PCM. En caso de requerirse conversiones a otros formatos de audio, estas se pueden realizar con el mismo utilitario.

Las grabaciones se efectuaron en una habitación cerrada, empleando un micrófono de condensador ubicado a unos 5 cm de la boca de las personas.

Para el desarrollo y prueba del algoritmo se empleó el paquete de software MATLAB, versión 6.5, para el entorno Windows. El MATLAB posee características que lo hacen muy

atractivo para el procesamiento de señales por sus potencialidades de cálculo, facilidad de programación y posibilidades de visualización. Se hizo uso especialmente del ToolBox de Procesamiento Digital de Señales.

También fueron utilizados algunos software libres, como el Colea y el Ecah, que brindan facilidades para analizar la señal del habla pues cuentan con herramientas que permiten seleccionar un segmento en específico dentro de la oración, así como, observar propiedades de parámetros importantes en el mismo, aspectos de gran interés en el análisis de las diferentes características asociadas al habla fluida y en la comprensión de las modificaciones articulatorias que ocurren, para llevar a cabo diferentes tipos de pruebas con el objetivo de validar la implementación.

2.2 Métodos

El método que se lleva a cabo para la realización de este trabajo consiste en tomar diferentes oraciones, aplicarle parte del algoritmo de Liu, poner las marcas en tiempo y comprobar su validez en el espectrograma. Para ello se describe el algoritmo general del segmentador seleccionado y sus modificaciones correspondientes y sobre esta base se programan las diferentes funciones.

2.2.1 Procesamiento general del algoritmo seleccionado

El procesamiento en general del algoritmo de Liu consiste en que un espectrograma es calculado y dividido en seis bandas de frecuencia y luego dos pasos de procesamiento (fino y grueso) son ejecutados. En cada paso una forma de onda de energía es construida en cada una de las seis bandas, la derivada de la energía es calculada y los picos en las derivadas son detectados. Estos picos representan tiempos de cambios espectrales abruptos en las seis bandas. En el procesamiento tipo-específico, los picos localizados son analizados para encontrar tres tipos de marcas: *glottis*, *sonorant*, *burst*. En la Figura 2.1 se representa el diagrama en bloques correspondiente.

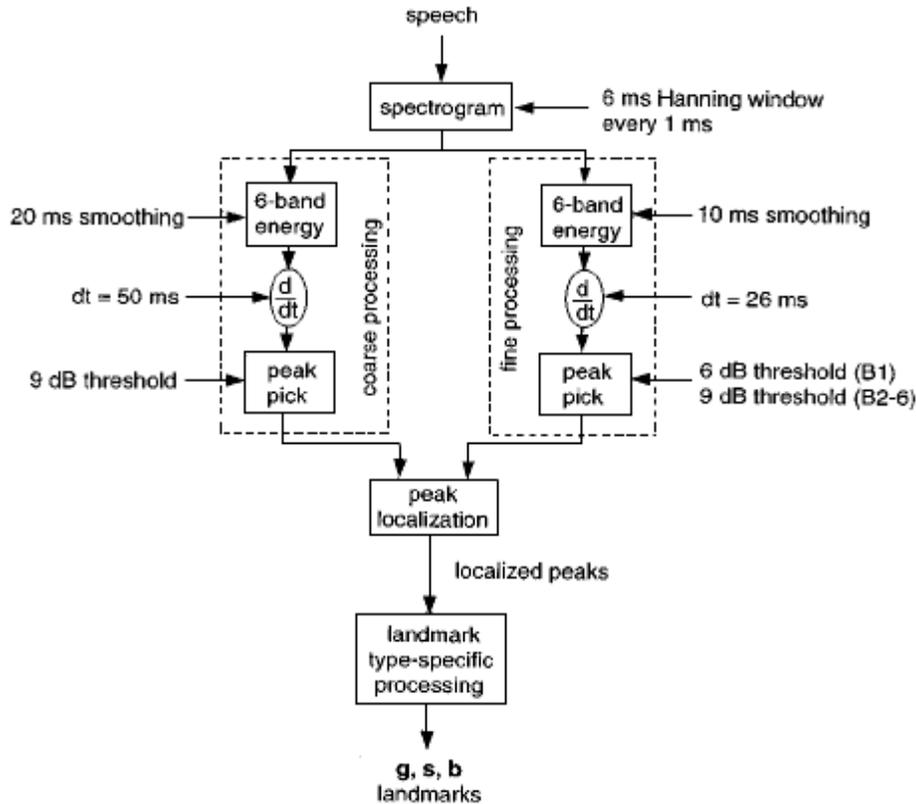


Figura 2.1. Algoritmo general según Liu.

Un espectrograma de banda ancha es calculado con una ventana Hanning de 6ms cada 1ms. Luego es aplicada la transformada discreta de Fourier con 512 puntos. El espacimiento entre los puntos para la transformada discreta de Fourier es de 31.2 Hz a fin de que las amplitudes de los picos espectrales en posteriores cálculos de energía puedan estar bien estimadas. Este espacio de 31.2 Hz entre los puntos se obtiene dividiendo la frecuencia de muestreo entre el número de puntos de la transformada. La alta tasa de segmentos permite que puedan ser monitoreados los cambios acústicos rápidos. Algunos cambios acústicos ocurren muy de prisa, particularmente los asociados con segmentos oclusivos. La pequeña ventana Hanning produce un espectro de banda ancha lo cual da amplia información espectral mientras los detalles armónicos son suprimidos.

El espectrograma resultante es dividido en las siguientes bandas de frecuencia:

Banda1: 0.0–0.4 kHz; banda2: 0.8–1.5 kHz; banda3: 1.2–2.0 kHz; banda4: 2.0–3.5 kHz; banda5: 3.5–5.0 kHz; banda6: 5.0–8.0 kHz.

BANDA 1: Monitorea la presencia o ausencia de vibración glotal. No se extiende por encima de 400Hz para reducir la oportunidad de recogida de ráfagas de energía en las bajas frecuencias. En un cierre de consonantes oclusivas sonoras o nasales, la energía en esta banda permanece fuerte porque la vibración glotal es continua. Sin embargo las prominencias espectrales por encima de F1 muestran una marcada disminución en la energía por el incremento de pérdidas acústicas.

BANDA 2-5: Cierres y liberaciones para consonantes sonoras. Aproximadamente en estas bandas se encuentran los rangos de frecuencia para las prominencias espectrales de consonantes sonoras. Los onsets y offsets de ruido de aspiración y fricación asociados con: consonantes oclusivas, fricativas y africativas también son encontrados en estas bandas. La energía de ruido estará en, al menos, una de estas cuatro bandas.

BANDA 2-3: Para segmentos de consonantes sonoras intervocálicas usualmente ocurre un cambio espectral grande en los rangos de frecuencia de 0.8 a 2kHz y estos cambios son a menudo debido a la introducción de un cero en la función de transferencia del tracto vocal en ese rango. En estas bandas no se garantiza el contenido de una prominencia espectral pero al menos una banda se espera que capture una prominencia espectral.

BANDA 6: Esta banda se extiende a lo largo de las frecuencias restantes hasta 8 kHz y es importante en la detección de silencio para oclusiones.

Después del cálculo del espectrograma se buscan los cambios de energía en las seis bandas usando una estrategia de dos pasos. Ambos pasos utilizan el mismo procesamiento, excepto que en el primer paso se utilizan parámetros más grandes con el fin de encontrar contornos generales de los cambios espectrales mientras que en el segundo paso los parámetros utilizados son más pequeños para localizar estos cambios de una manera más precisa. En la Figura 2.2 se muestran los resultados obtenidos por Liu para la banda 1 del espectrograma de la oración *The money is coming today*.

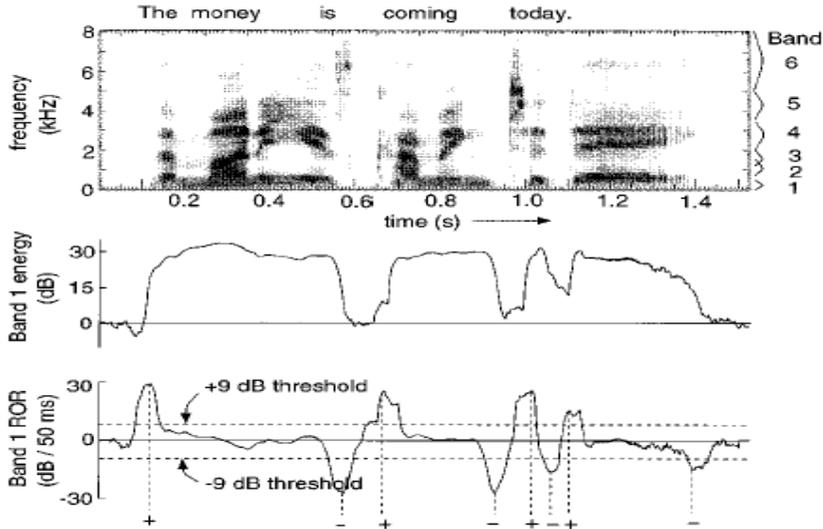


Figura 2.2. Resultados para la banda 1 según Liu.

En el **paso de procesamiento grueso** primeramente una forma de onda de la energía es calculada para cada una de las seis bandas, la cual se realiza para resolver cambios acústicos abruptos debido a los cambios bruscos en las amplitudes de los formantes de las frecuencias, pero sin tener en cuenta variaciones de los pulsos glotales y fluctuaciones de ruido.

Para disminuir las características no deseadas un promedio de 20ms de la magnitud cuadrada del espectrograma, centrado sobre el tiempo de interés, es calculado cada 1ms. En cada una de las bandas el máximo valor en cada instante de tiempo es tomado para representar la energía en esa banda y la misma es dada en dB. Como una banda abarca exactamente una prominencia espectral se selecciona el máximo valor de energía en la banda como una función de tiempo y esto es lo mismo que guiarse por la amplitud de prominencia espectral en el tiempo.

Una vez que la energía en todas las bandas es calculada, se buscan los cambios relacionados con las variaciones de energía (ROR). La forma de onda del ROR (rate of rise) indica la rapidez con que se altera la energía en cada banda.

La primera diferencia es calculada cada 1ms usando un tiempo de paso de 50ms centrado sobre el tiempo de interés. El paso de 50ms es seleccionado con el objetivo de extenderse a lo largo de transiciones de energía para cierres y liberaciones abruptos, incluyendo cierres

de consonantes oclusivas sonoras, teniendo en cuenta la ventana Hanning de 6ms y los 20ms de suavizamiento.

Los picos positivos y negativos en la forma de onda del ROR son los puntos de cambios acústicos abruptos en una banda. En un algoritmo de Mermelstein (1975a) para elegir picos se seleccionan aquellos cuyo valor absoluto es mayor que 9dB. El umbral de 9dB esta dado por la diferencia en la amplitud de la fuente glotal entre una consonante oclusiva y un segmento vocálico vecino.

En el **paso de procesamiento fino** algunos valores de parámetros son modificados para localizar cambios de energía en el tiempo. Un intervalo de 10ms de suavizamiento es utilizado en el espectrograma, mientras que el tiempo de paso seleccionado para la primera diferencia es de 26ms y el umbral para los picos de la banda 1 es reducido de 9dB a 6dB debido a la reducción del tiempo de paso. Los umbrales para los picos de las bandas 2, 3, 4, 5 y 6 son mantenidos en 9dB para impedir picos espurios.

Los picos del ROR obtenidos tanto en el paso grueso como en el fino pasan al bloque de localización de picos. Aquí los picos del paso grueso son utilizados como guía para encontrar los picos del ROR correspondientes al paso fino. Dentro de aproximadamente 30ms de un pico grueso, el pico más grande en términos absolutos en el paso fino es seleccionado como el pico localizado.

Los picos localizados son las entradas de la etapa de procesamiento de marcas especificadas por tiempo.

Tipos de marcas

Detector G (glottis)

Una marca **g** define un tiempo de comienzo o fin de vibración de las cuerdas vocales. Los picos localizados en el ROR de la banda1 del procesamiento general son todos inicialmente candidatos para marcas **g**. Estos candidatos deben pasar una serie de criterios. Un pico positivo indica el inicio de vibración glotal y un pico negativo indica la terminación de vibración glotal. Cuando la vibración glotal comienza debe terminar un tiempo después. Así cada pico positivo debería ser seguido por un pico negativo. Los picos son insertados

donde sea necesario para satisfacer estas condiciones. El punto de inserción está guiado por la forma del contorno de la energía de la banda 1.

Detector S (sonorant)

Una marca *s* es causada por el cierre o liberación de una consonante nasal o líquida. Un *s* negativo designa un cierre y uno positivo designa una liberación. Como en el tracto vocal existe una constricción para consonantes sonoras, la energía en el rango de F2 a F1 disminuye y en una liberación esta energía aumenta. Si la constricción es lo suficientemente apretada y ocurre de forma rápida, entonces la energía cambiará rápida y simultáneamente en todas las bandas. Durante el intervalo de constricción, para una consonante nasal o líquida, un articulador primario ha hecho un cierre completo, las cuerdas vocales continúan vibrando y la forma del tracto vocal es relativamente constante. Por lo tanto el espectro debería permanecer relativamente estable especialmente en bajas frecuencias.

Para encontrar marcas *s*, solo en regiones donde existe sonoridad, nos ubicamos en aquellas regiones que se encuentran entre un *g* positivo a la izquierda y un *g* negativo a la derecha. Dentro de una región sonora, son agrupados cualquier pico con el mismo signo en las bandas de la 2 a la 5. El pico absoluto más grande en cada grupo es el llamado pivot y es un candidato para una marca *s*. En la Figura 2.3 se muestra un ejemplo de los resultados alcanzados por el algoritmo de Liu para marcas *sonorant*.

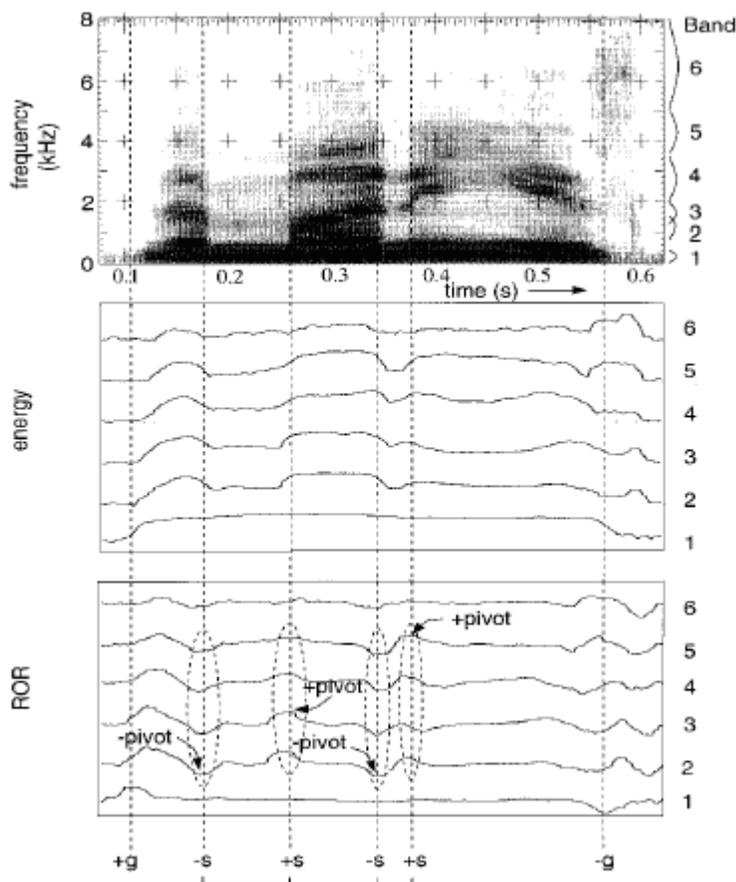


Figura 2.3. Detección de marcas sonorant según el algoritmo de Liu.

Generalmente consonantes sonoras oclusivas (e.g., [d, v]) son a menudo difíciles de detectar con el detector de marcas **g** porque en la energía de la banda 1 no ocurren cambios suficientes. Ellos, sin embargo, usualmente muestran cambios de energía más claros en altas frecuencias. Si escapan del detector **g**, entonces pueden ser detectados por una variación del detector **s**. En esta variación los pivot son encontrados y las alteraciones en las altas frecuencias son requeridas como antes, sin embargo, la energía en las bajas frecuencias, en medio de un cierre y liberación del obstructor, no debe ser de un estado estable.

Detector B (burst)

Una marca **b** positiva significa una africación o una ráfaga de fin de aspiración. Los correlativos acústicos para una marca **b** positiva son un intervalo de silencio seguido por un incremento bien definido de la energía en las altas frecuencias. A partir de que marcas **b** solo pueden ocurrir durante regiones sin vibración glotal, aquellas regiones delimitadas por

un **g** negativo a la izquierda y un **g** positivo a la derecha son tomadas por este detector. Primeramente, opivots (para pivot de consonantes oclusivas sordas) son encontrados de una manera análoga. Un opivot es candidato para una marca **b**. Después el silencio es medido alrededor de un opivot. Para un opivot positivo, un intervalo de silencio debe existir a la izquierda. Este período de silencio es medido con la energía de las bandas de la 3 a la 6, usando los niveles de energía en cada banda como referencia. Una marca **b** negativa señala la terminación de una fricación o ruido de aspiración. Los correlativos acústicos para una marca **b** negativa son una disminución de energía en las altas frecuencias seguida por un intervalo de silencio. Este silencio es medido usando todas las bandas, incluyendo las bandas 1 y 2.

En la Figura 2.4 se muestra el resultado alcanzado por el detector Burst del algoritmo anterior; las marcas **b** señaladas definen un intervalo de silencio encontrado en un segmento sordo definido por las marcas **g**.

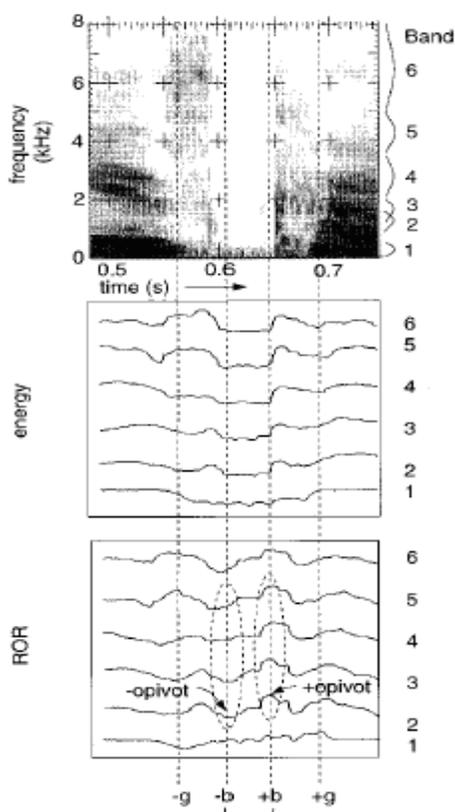


Figura 2.4. Detección de marcas Burst según el algoritmo de Liu.

2.2.2 Modificaciones realizadas e implementación

Para la implementación del segmentador automático que tiene como propósito este trabajo se realizaron una serie de modificaciones al algoritmo de Liu. Se escogió solamente el procesamiento fino porque este incluye los cambios espectrales encontrados por el procesamiento grueso y de una forma más precisa. También muestra las variaciones de la señal del habla en las distintas bandas que no son encontradas por el paso grueso pues este último encuentra contornos generales de los cambios espectrales. Además de que el procesamiento grueso pasa por alto características importantes no es necesario para encontrar los picos correspondientes en la misma banda del paso fino pues este consta de sus propios parámetros y se puede llegar por esta vía al resultado esperado, como se describe en el apartado anterior. Es importante señalar que los valores de los parámetros del paso seleccionado no fueron alterados, es decir, el cambio en esta primera etapa solo radica en utilizar un procesamiento y no dos.

Basada en las marcas analizadas anteriormente se realizaron ajustes para localizar los cambios espectrales de interés en una señal del habla dada, utilizando solamente parámetros finos, pero siguiendo prácticamente el mismo procedimiento.

- **g** (glottis). Marca instantes de tiempo en los cuales comienza y termina la vibración de las cuerdas vocales, es decir, segmentos sonoros y sordos en regiones del espectrograma, como por ejemplo, los asociados a cierres de consonantes fricativas, transiciones de consonantes nasales a fricativas, consonantes oclusivas, fin y comienzo de palabras, entre otras variaciones que ocurren en el habla fluida.
- **s** (sonorant). Cambios espectrales dentro de regiones de sonoridad, medidos en las bandas 2 y 3, los cuales están asociados generalmente a consonantes intervocálicas.
- **b** (burst). Variaciones espectrales de importancia en regiones sordas, medidas en las bandas de la 2 a la 6, como pueden ser intervalos de silencios que incluyen pausas, consonantes oclusivas y fricativas o un fin de aspiración.

Las funciones implementadas para cada uno de los bloques fueron las siguientes:

Teniendo los datos de la señal de entrada y la frecuencia de muestreo, la función *enbandapf* divide la señal en tramas de 6ms cada 1ms y a este resultado le calcula la transformada discreta de Fourier para 512. Además devuelve las seis bandas de frecuencia.

Energiapf: Devuelve los valores de energía de cada una de las bandas, calculados para un intervalo de 10ms cada 1ms (Ver la Figura A.1 del Anexo A).

Derivpf: Mediante esta función se obtienen los valores de los puntos donde ocurren cambios acústicos en las diferentes bandas de energía, calculados para un tiempo de paso de 26ms cada 1ms (Ver la Figura A.2 del Anexo A).

Umbralpf: Devuelve las posiciones y amplitudes de los picos positivos y negativos mayores que 9dB (6dB para la banda1) y menores que -9dB (-6dB para la banda1) respectivamente y los muestra gráficamente (Ver la Figura A.3 del Anexo A).

Las funciones anteriores están recogidas en el fichero **edpf.m** (Ver el programa correspondiente en el Anexo B)

Para mostrar los distintos tipos de marcas en el espectrograma de una oración determinada se debe seguir la siguiente secuencia.

Secuenciapf: Organiza los picos localizados por la función *umbralpf* en pares, es decir, de forma tal que después de un pico positivo le siga uno negativo y viceversa. Además devuelve las localizaciones de estos picos. (Ver la Figura A.4 del Anexo A y el programa correspondiente en el Anexo B).

Glottispf: Con los resultados anteriores devuelve las marcas **g** y las representa en el gráfico obtenido por la función anterior (Ver la Figura A.5 del Anexo A y el programa correspondiente en el Anexo B).

Sonorantpf: Localiza los intervalos entre un pico negativo y uno positivo dentro de una región sonora medidos en las bandas 2 y 3 devolviendo las localizaciones de las marcas **s** (Ver la Figura A.6 del Anexo A y el programa correspondiente en el Anexo B).

Burstpf: Localiza los intervalos entre un pico negativo y uno positivo dentro de una región sorda medidos desde la banda 2 a la 6 teniendo en cuenta la banda 1y devuelve las localizaciones de las marcas **b** (Ver la Figura A.7 del Anexo A y el programa correspondiente en el Anexo B).

Markspecgpf: Muestra tres espectrogramas donde se ilustran los cambios espectrales localizados por cada tipo de marca y las diferencias en tiempo entre las marcas negativas y

positivas en cada caso (Ver las Figuras A.8, A.9 y A.10 del Anexo A y el programa correspondiente en el Anexo B).

Conclusiones parciales

El algoritmo implementado presenta un gran potencial para el análisis del habla fluida. El mismo destaca una parte importante de los cambios espectrales que ocurren en una oración o en un segmento del habla dado. A partir de las modificaciones realizadas al algoritmo de Liu, se facilita la implementación del mismo porque se reduce el número de pasos a seguir y además se obtienen buenos resultados para el objetivo que se persigue en este trabajo.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En el presente capítulo se muestra el desempeño del algoritmo programado.

Teniendo en cuenta que el algoritmo original fue desarrollado a partir del análisis de espectrogramas, se han tomado un conjunto de oraciones, algunas en Inglés y otras en Español, para encontrar en las mismas las marcas espectrales de interés. Las marcas se sitúan sobre el espectrograma de la oración de la cual se extrajeron para, de esa forma, validar el funcionamiento del programa.

3.1 Validez y funcionalidad de las marcas

A continuación se muestran algunos de los resultados obtenidos después de realizada la programación correspondiente de las funciones explicadas en el capítulo anterior. Para una mejor comprensión y visibilidad de los resultados se escogió un espectrograma para cada tipo de marca pues en ocasiones, como veremos en los ejemplos siguientes, coinciden dos tipos de marcas para mostrar una variación acústica determinada. Es importante señalar la información que brindan las marcas representadas en las figuras, por lo cual, a continuación se presenta la leyenda de las mismas.

Leyenda de las marcas

+g (verde) Incremento de sonoridad (Inicio de la vibración).

-g (rojo) Decremento de la sonoridad (Fin de la vibración).

+b (azul) Incremento de la energía (Fin del silencio).

-b (cian) Decremento de la energía (Inicio del silencio).

+s (magenta) Incremento de la energía en las bandas intermedias (Fin de consonante intervocálica).

-s (amarillo) Decremento de la energía en las bandas intermedias (Inicio de consonante intervocálica).

Primeramente se muestran dos ejemplos de oraciones en Inglés. En las frases se analizan los cambios espectrales delimitados por las marcas y que confirman el funcionamiento apropiado de las mismas.

1. You want to know all about my grandfather.

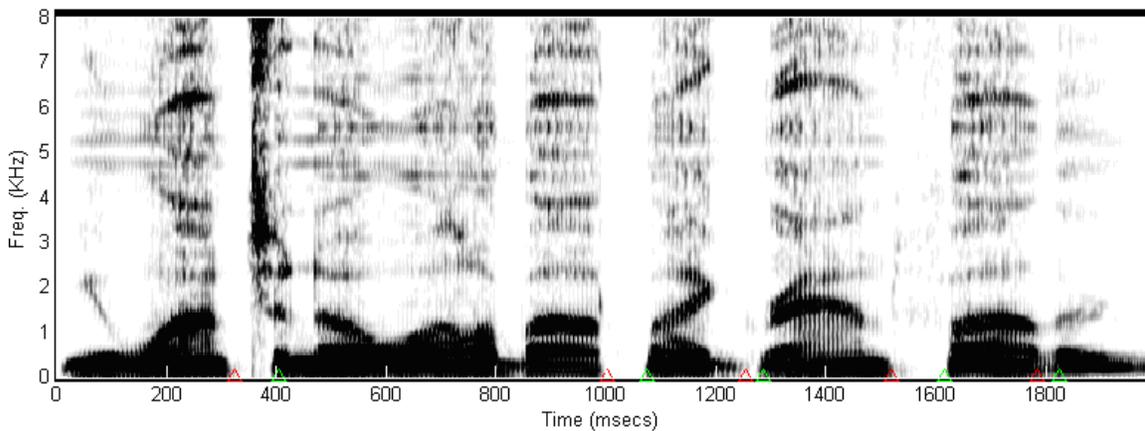


Figura 3.1. Espectrograma con las marcas correspondientes a segmentos sonoros y sordos.

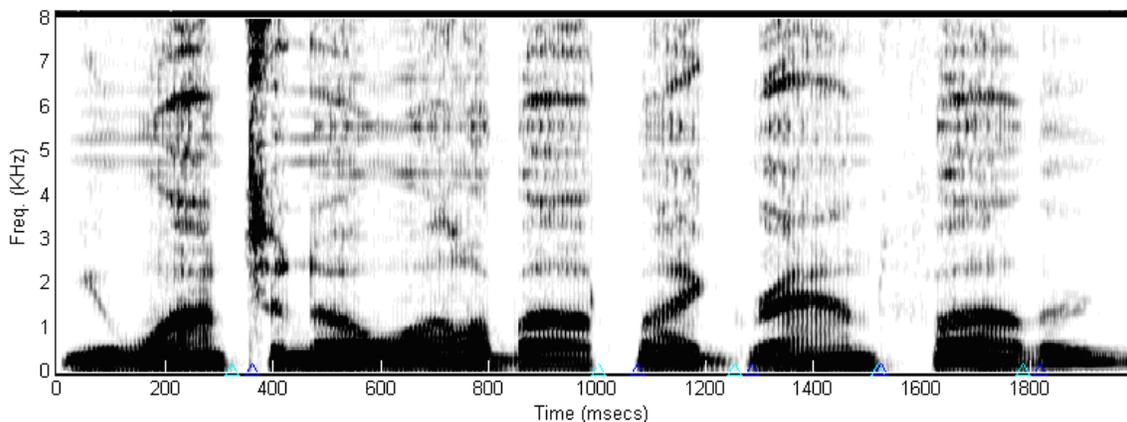


Figura 3.2. Espectrograma con las marcas correspondientes a segmentos de silencio.

Como mencionamos anteriormente, en la Figura 3.2 algunas marcas coinciden con las obtenidas en la Figura 3.1 y esto se debe a que cuando hay ausencia de vibración de las cuerdas vocales, puede existir solo un cambio espectral, como ejemplo se encuentra el

segundo, tercero y quinto par de marcas mostrados en la Figura 3.2; o varios cambios espectrales como son los encontrados en el primer intervalo sordo, delimitado por una marca **g** negativa a la izquierda y una marca **g** positiva a la derecha, como se ilustra en la Figura 3.1, que presenta un silencio seguido por un intervalo de tiempo ruidoso antes de que comiencen a vibrar las cuerdas vocales. Se representa en este caso, una consonante oclusiva sorda.

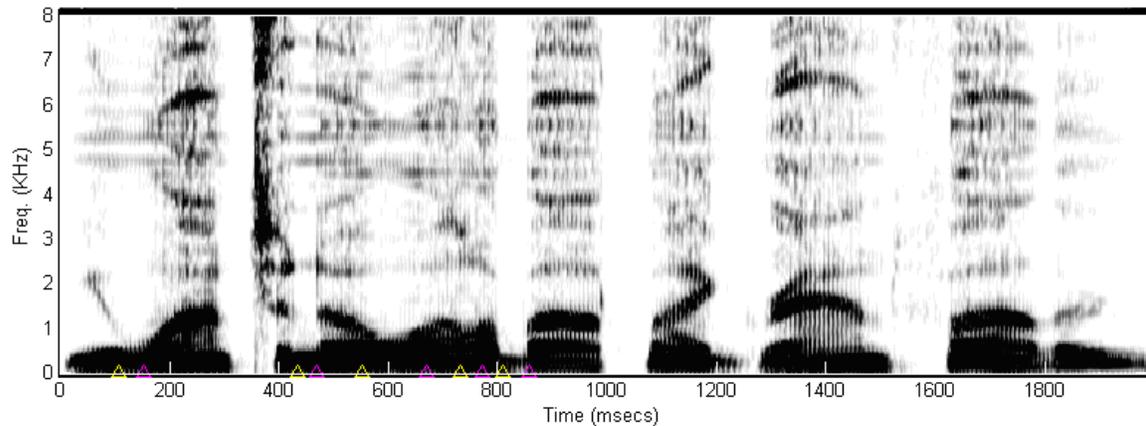


Figura 3.3. Espectrograma con las marcas correspondientes a segmentos de consonantes intervocálicas.

En el caso de la Figura 3.3 no existe correspondencia alguna con otros tipos de marcas. Estos cambios capturados son asociados fundamentalmente a consonantes sonoras intervocálicas como *n*, *l*, y *b* en este caso.

2. *The money is coming today.*

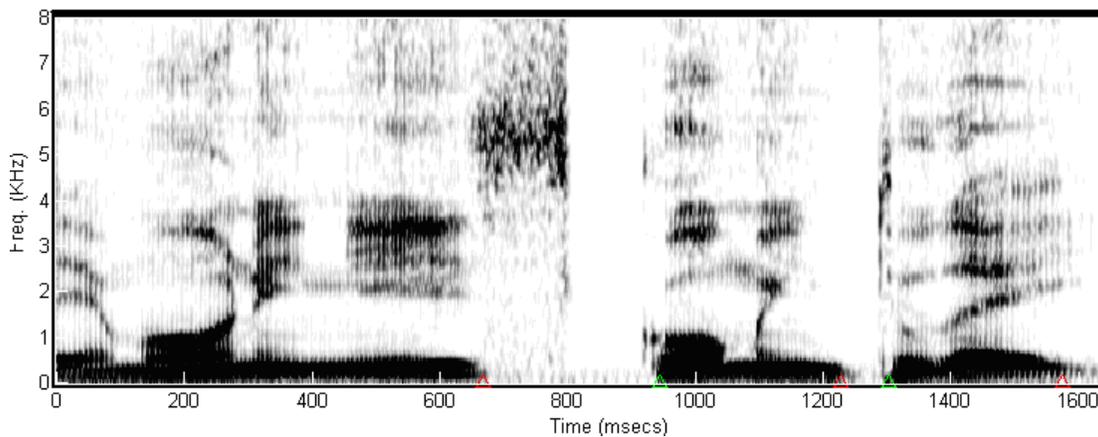


Figura 3.4. Espectrograma con las marcas correspondientes a segmentos sonoros y sordos.

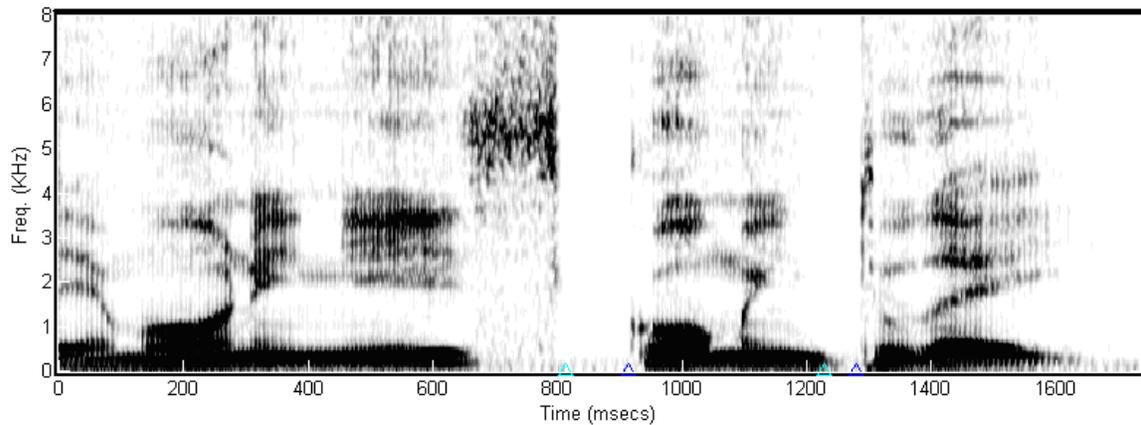


Figura 3.5. Espectrograma con las marcas correspondientes a segmentos de silencio.

En las Figuras 3.4 y 3.5 se pueden comprobar los resultados analizados en la oración anterior. En este ejemplo la presencia de un segmento sordo delimitado entre la primera y segunda marca de la Figura 3.4 incluye un segmento sin sonoridad con variaciones en las altas frecuencias, correspondiente a una consonante fricativa, seguido por un intervalo de silencio definido por las fronteras uno y dos de la Figura 3.5. Este silencio es el que conforma el tiempo de oclusión de la /k/ en *coming*.

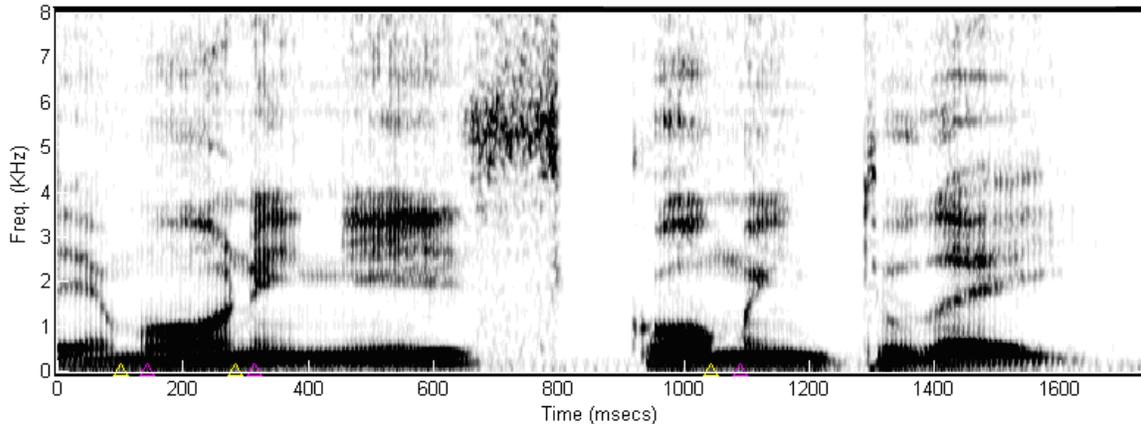


Figura 3.6. Espectrograma con las marcas correspondientes a segmentos de consonantes intervocálicas.

En la Figura 3.6 se observa que las consonantes nasales m y n han sido detectadas adecuadamente a pesar de estar en posición intervocálica y su corta duración. Lo cual constituye una de las fortalezas de esta implementación.

Como se puede observar en estos dos ejemplos se obtienen, de forma automática y con una buena exactitud, cambios espectrales de interés.

En los espectrogramas que se muestran a continuación se comprueban los mismos resultados pero con frases en Español. Esto valida el funcionamiento del método independientemente del idioma.

3. *Esto es una prueba para la tesis de Adilson.*

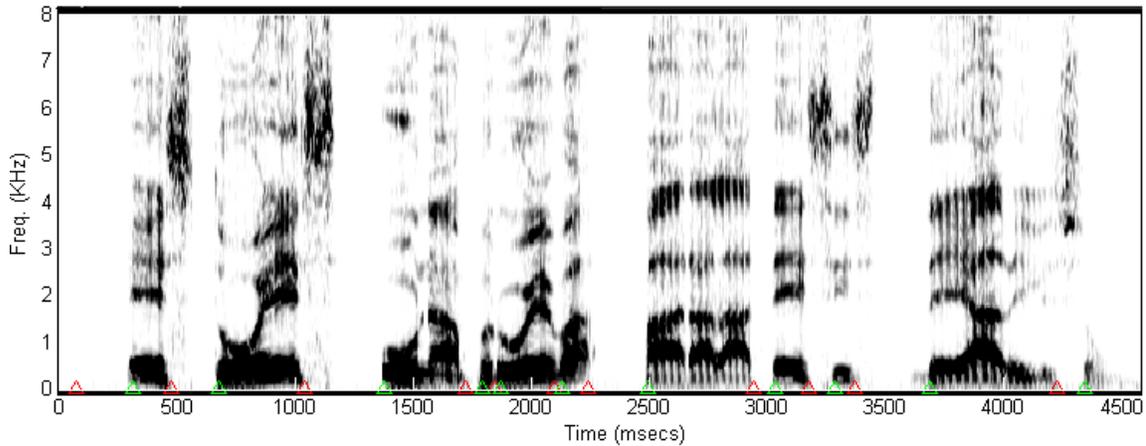


Figura 3.7. Espectrograma con las marcas correspondientes a segmentos sonoros y sordos.

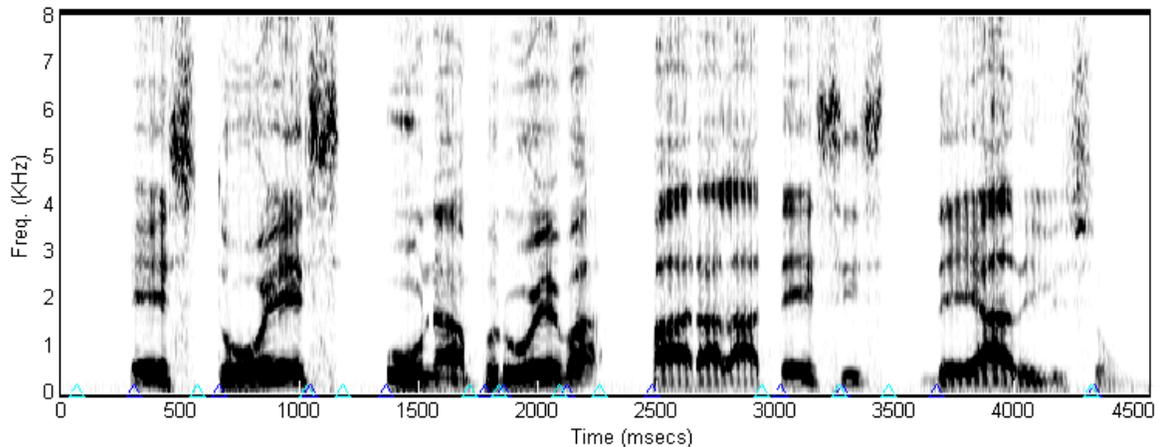


Figura 3.8. Espectrograma con las marcas correspondientes a segmentos de silencio.

En las Figuras 3.7 y 3.8 se obtienen buenos resultados y se manifiestan las mismas características expuestas en las oraciones en Inglés, es decir, independientemente en qué idioma, de los analizados en el trabajo, sea pronunciada la oración los resultados serán los mismos, ya que se obtendrán, a través de las funciones desarrolladas, los cambios espectrales asociados a las características distintivas.

En las distintas comprobaciones realizadas se puede ver que los intervalos de tiempo asociados a las marcas **g** contienen el mayor peso para detectar las variaciones acústicas

porque debido a sus características agrupan la mayor cantidad de categorías fonéticas y son una guía para encontrar los demás tipos de marcas. De hecho las marcas **g** señalan el inicio y fin de un núcleo vocálico.

Para cada par de marcas las funciones implementadas devuelven además, el valor del tiempo entre ellas. Este resultado permite a los especialistas analizar de forma detallada cada uno de los cambios ocurridos.

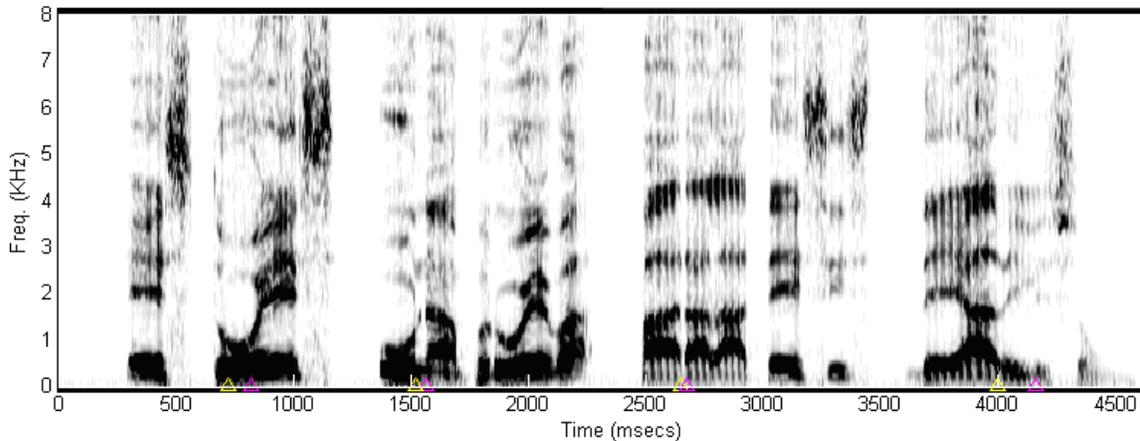


Figura 3.9. Espectrograma con las marcas correspondientes a segmentos de consonantes intervocálicas.

En la Figura 3.9 puede apreciarse que las marcas señalan la ocurrencia de consonantes intervocálicas como n, r y d (pares de marcas dos, tres y cuatro respectivamente) tanto como uniones en el caso de oe en *esto es*, todo lo cual da lugar a cambios espectrales notables.

3.2 Observaciones en el habla fluida

A continuación se brindan los resultados obtenidos para dos personas diferentes pronunciando la misma oración ya que en general esas diferencias ocurren.

4. Veremos si vamos mañana a la playa.

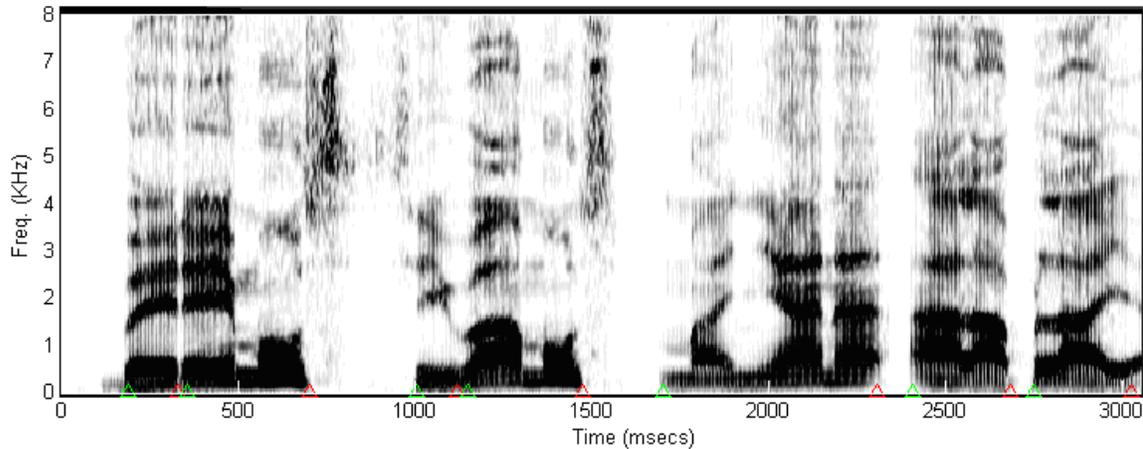


Figura 3.10. Espectrograma de la frase, mencionada por la primera persona, con las marcas correspondientes a segmentos sonoros y sordos.

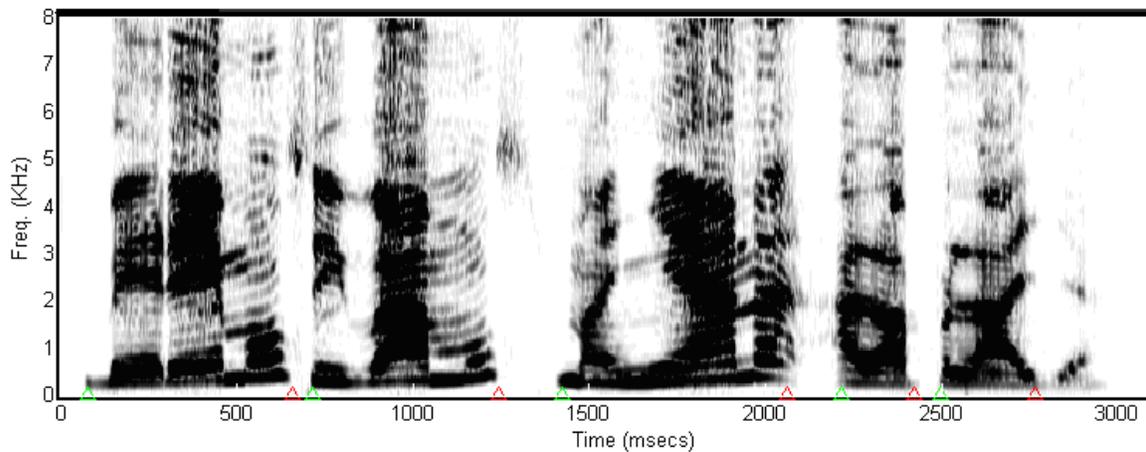


Figura 3.11. Espectrograma de la frase, mencionada por la segunda persona, con las marcas correspondientes a segmentos sonoros y sordos.

Como se puede observar en las Figuras 3.10 y 3.11 los espectrogramas no son completamente iguales por lo que las marcas en una y otra figura no coinciden en su totalidad. Estos factores están determinados por las diferencias en el habla de las personas en cuanto a los diferentes aspectos de la prosodia. Pueden existir para una misma oración, hablada por personas diferentes, cambios acústicos que no se correspondan. Por ejemplo en la Figura 3.10 las marcas dos y tres delimitan un intervalo de fin y comienzo de sonoridad asociado a una transición de vocal a consonante líquida, mientras que en la Figura 3.11 este cambio no es detectado. Lo mismo sucede con las marcas seis y siete de la Figura 3.10 que

señalan la presencia de una consonante oclusiva sonora después de una sílaba, para este caso la parte de la oración es: *si vamos*, en la Figura 3.11 este cambio no se encuentra marcado ya que el sonido oclusivo se vuelve fricativo.

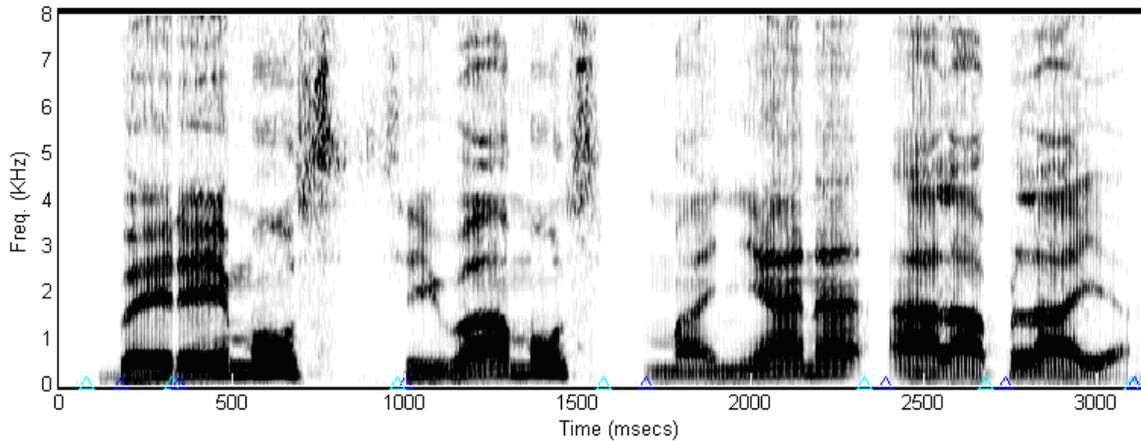


Figura 3.12. Espectrograma de la frase, mencionada por la primera persona, con las marcas correspondientes a segmentos de silencio.

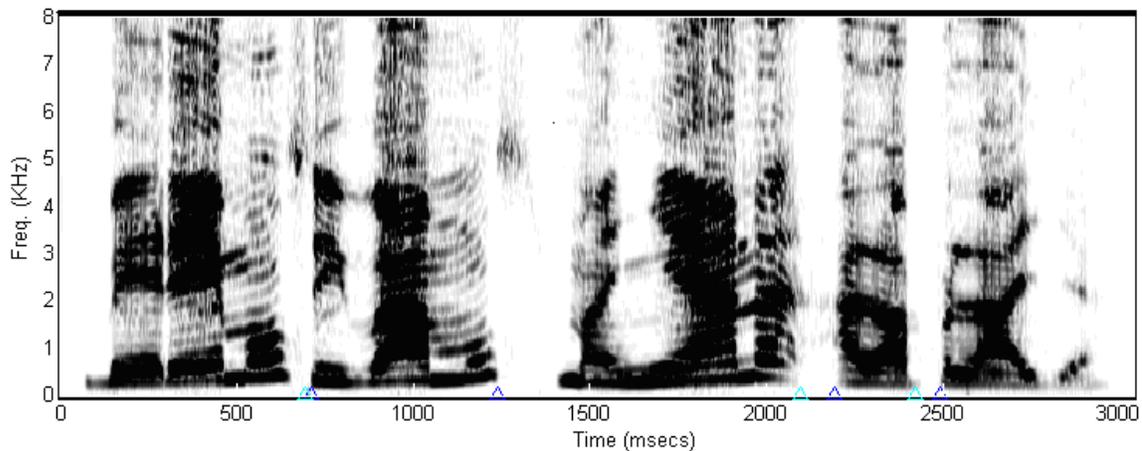


Figura 3.13. Espectrograma de la frase, mencionada por la segunda persona, con las marcas correspondientes a segmentos de silencio.

En el caso de los segmentos de silencio ocurre algo semejante, como se muestra en las Figuras 3.12 y 3.13. Otras de las características que influyen en estos factores son los procesos de asimilación y co-articulación, los cuales, están más marcados en unas personas que en otras. Estas modificaciones que influyen en el habla fluida provocan que intervalos de cambios espectrales, delimitados por un tipo de marca en el habla de una persona determinada, sean encontrados por otros tipos de marcas para otras personas. Esto se puede

observar en la Figura 3.12 entre las marcas tres y cuatro, donde existe un pequeño silencio, mientras que en la Figura 3.15, para el caso de la mujer que se tomó como ejemplo, este cambio es señalado como consecuencia de una consonante intervocálica debido a modificaciones en los fonemas y sonidos vecinos.

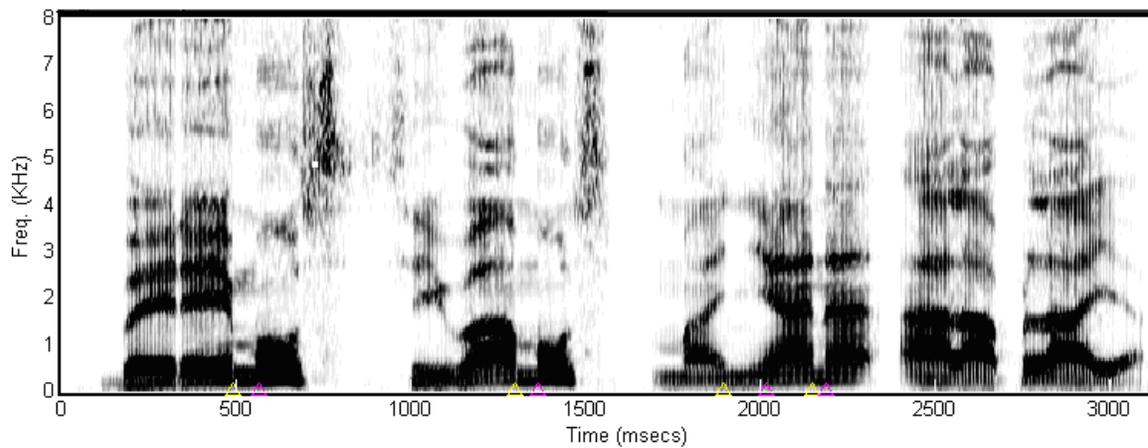


Figura 3.14. Espectrograma de la frase, mencionada por la primera persona, con las marcas correspondientes a segmentos de consonantes intervocálicas.

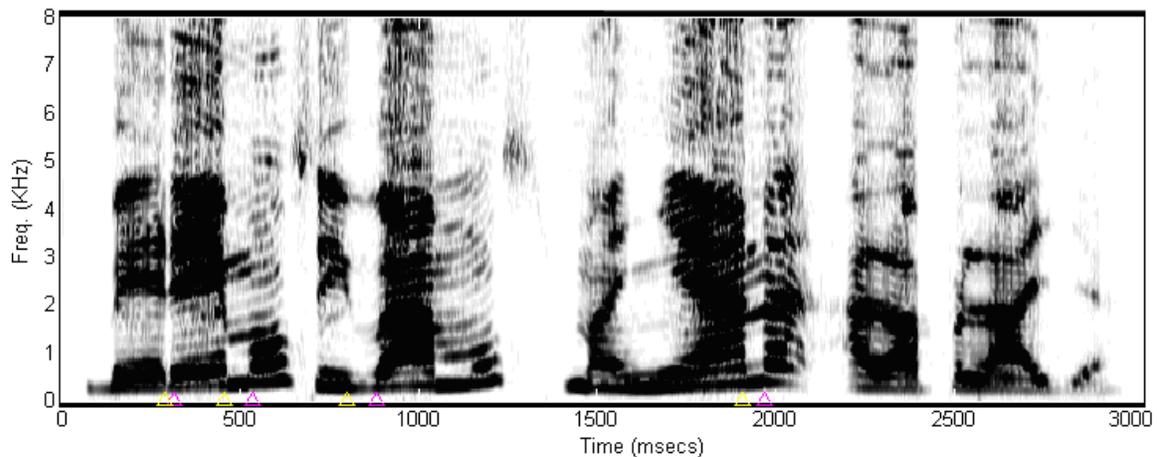


Figura 3.15. Espectrograma de la frase, mencionada por la segunda persona, con las marcas correspondientes a segmentos de consonantes intervocálicas.

Las marcas cinco y seis de la Figura 3.15 denotan un cambio espectral de una consonante intervocálica debido a que, en este caso, ocurre una modificación a las articulaciones, pues en esta misma parte de la oración (*si vamos*) en la Figura 3.14 este cambio no es encontrado, como se mencionó con anterioridad.

Conclusiones parciales

A través de los ejemplos mostrados, se ha constatado que la implementación realizada basada en el algoritmo de Liu, destaca los cambios espectrales que ocurren durante el habla fluida y que pueden considerarse como puntos de inflexión a tener en cuenta en un análisis prosódico posterior.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

La elaboración del presente trabajo ha permitido arribar a las siguientes conclusiones y se han obtenido resultados que merecen ser destacados:

- El estudio realizado de diferentes tipos de segmentadores validados en la literatura teniendo en cuenta sus parámetros y características individuales, así como, las definiciones de los distintos aspectos de la prosodia sistematizados en el trabajo, permite a los especialistas interesados conocer qué tipo de segmentador utilizar para una aplicación en particular.
- Los ejemplos citados incluyen comentarios del autor de este trabajo, los mismos aportan elementos que pueden contribuir a la elección de uno en particular y están fundamentados en la bibliografía consultada.
- Las modificaciones realizadas al algoritmo seleccionado han permitido obtener buenos resultados y disminuyen en gran medida la multiplicidad de la implementación.
- De las pruebas realizadas se puede concluir que, aunque pudieran buscarse otros cambios en la señal del habla combinando las diferentes marcas, se consideran de un gran significado los procedimientos ya desarrollados porque recogen una parte importante de las características del habla fluida.

Recomendaciones

- Partiendo de las funciones implementadas, completar los algoritmos para destacar los aspectos prosódicos a partir de los cambios acústicos capturados en la señal del habla.
- Analizar el segmentador de Salomon con el objetivo de obtener resultados semejantes y/o complementarios aún en el caso de que la señal esté degradada.

REFERENCIAS BIBLIOGRÁFICAS

- Bitar, N. (1997). Acoustic analysis and modeling of speech based on phonetic features, Boston University, Boston.
- Bronstein, A. J. (1960). *The pronunciation of American English*. Appleton-Century-Crofts, New York.
- Cristal, D. (1982). *Clinical linguistics*. Springer-Verlag, Viena.
- Cruttenden, A. (1986). *Intonation*. Cambridge University Press, Cambridge.
- Epstein, M. (2002). Voice Quality and Prosody in English, Tesis de doctorado, Universidad de California, Los Angeles.
- Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjr/ in American English, *J. Acoust. Soc. Am.*, 92, 736–757.
- Ferrer, C. (2005). Cuantificación de la voz para aplicaciones medicas, Tesis de doctorado, Universidad central de Las Villas, Santa Clara.
- Flege, J. E. (1988). The development of skill in producing word-final English stops: Kinematic parameters, *Journal of the Acoustical Society of America*, 84, 1639-1652.
- Gili Gaya, S. (1993). *Estudios sobre el ritmo*. Istmo, Madrid.
- Hess, W. (1975). A Pitch-Synchronous Digital Feature Extraction System for Phonemic Recognition of Speech, *IEEE Trans. Acoust. Speech Signal Process*, 25.
- Ingram, D. (1976). *Phonological disability in children*. Elsevier, New York.
- Jakobson, R., Fant, G., and Halle, M. (1952). Preliminaries to speech analysis, MIT Acoustics Laboratory.

- Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory, *Journal of Phonetics*, 14, 29-59.
- Kent, R. and Bale, M. (2000). *Voice Quality Measurement*. Singular.
- Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich, New York.
- Laver, J. (1991). *The gift of speech*, Edinburgh University Press, Edinburgh, UK.
- Lehiste, I. (1970). *Suprasegmental*. MIT Press, Cambridge, MA.
- Liu, S. (1996). Landmark detection for distinctive feature-based speech recognition. Research Lab of Electronics and Department of Electrical Engineering and Computer Science, Massachusetts of Technology, Cambridge, Massachusetts, 36-511.
- MacKay, I. (1987). *Phonetics: The science of speech production*. Little Brown, Boston.
- Memelstein, P. (1975). Phonetic-Context Controlled Strategy for Segmentation and Phonetic Labeling of speech, *IEEE Trans. Acoust. Speech Signal Process*, 23.
- Mermelstein, P. (1975a). Automatic segmentation of speech into syllabic Units, *J. Acoust. Soc. Am.*, 58, 880–883.
- Ohde, R., & Sharf, D. (1992). *Phonetic analysis of normal and abnormal speech*. Macmillan, New York.
- Rabiner, L. R. & Sambur, M. R. (1975). Some Preliminary Experiments in the Recognition of Connected Digits, *IEEE Trans. Acoust. Speech Signal Process*, 24, 170-171.
- Salomon, A. (2004). Detection of speech landmarks: Use of temporal information, Electrical and Computer Engineering Department, University of Maryland, A.V Williams Building, College Park, Maryland.
- Shriberg, L. D., & Kent, R. D. (1982). *Clinical phonetics*. John Wiley & Sons, New York.
- Smith, B. L., & McLean-Muse, A. (1986). Articulatory movement characteristics of labial consonant productions by children and adults, *Journal of the Acoustical Society of America*, 80, 1321-1328.

Smith, B. L., & McLean-Muse, A. (1987). Effects of rate and bite block manipulations on kinematic characteristics of children's speech, *Journal of the Acoustical Society of America*, 81, 747-754.

Susand, S. and Rabiner, L. (2006). On the relation between Maximum Spectral Transitions and phone boundaries, In Proc. Interspeech, Pittsburgh, Pennsylvania, 17-21.

Tiffany, W. R., & Carrell, J. (1977). *Phonetics: Theory and application*. McGraw-Hill, New York.

Viikki, O. (2001). Noise robust ASR, *Speech Commun*, 34.

Xin, L. (2005). Introduction to Digital Speech Processing. [En línea]. Disponible en: <http://www.csee.wvu.edu/~xinl/courses/ee493q>. Visitado el: 6/6/08

ANEXOS

Anexo A Funciones implementadas

A continuación se muestran los resultados que se alcanzan por cada una de las funciones en el orden que se efectúan (Ver Figuras A.1, A.2, A.3, A.4, A.5, A.6, A.7, A.8, A.9 y A.10). La oración analizada para mostrar estos ejemplos fue la siguiente:

Estamos en la tesis

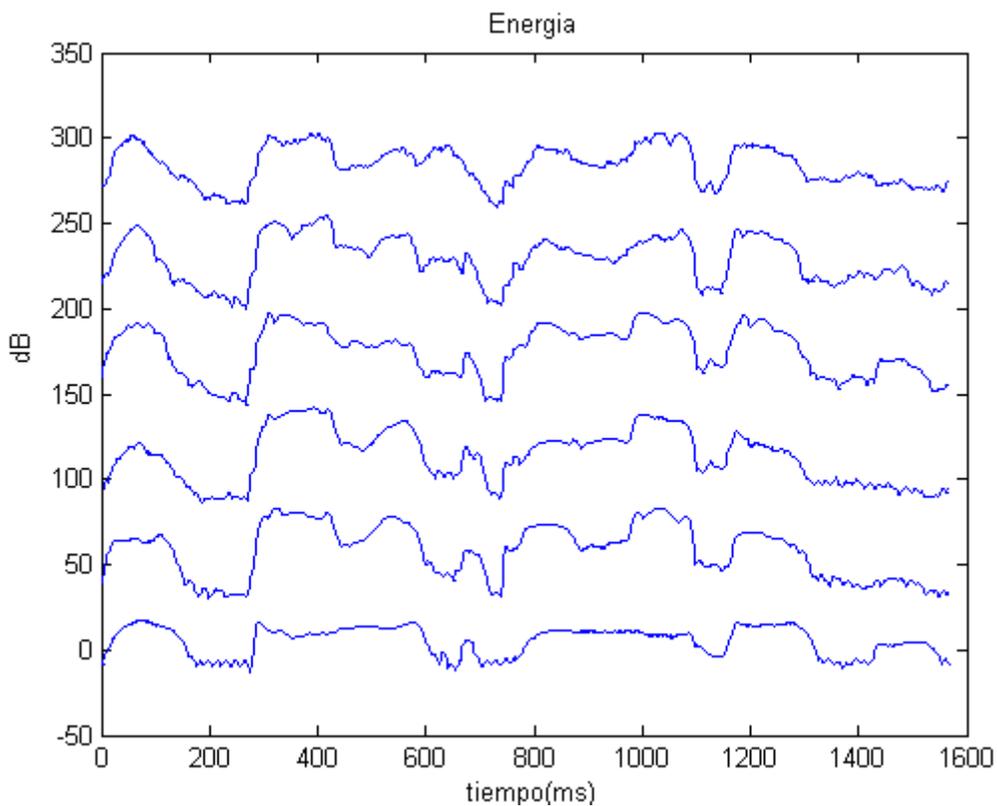


Figura A.1. Variaciones de energía de las seis bandas (*energiapf*).

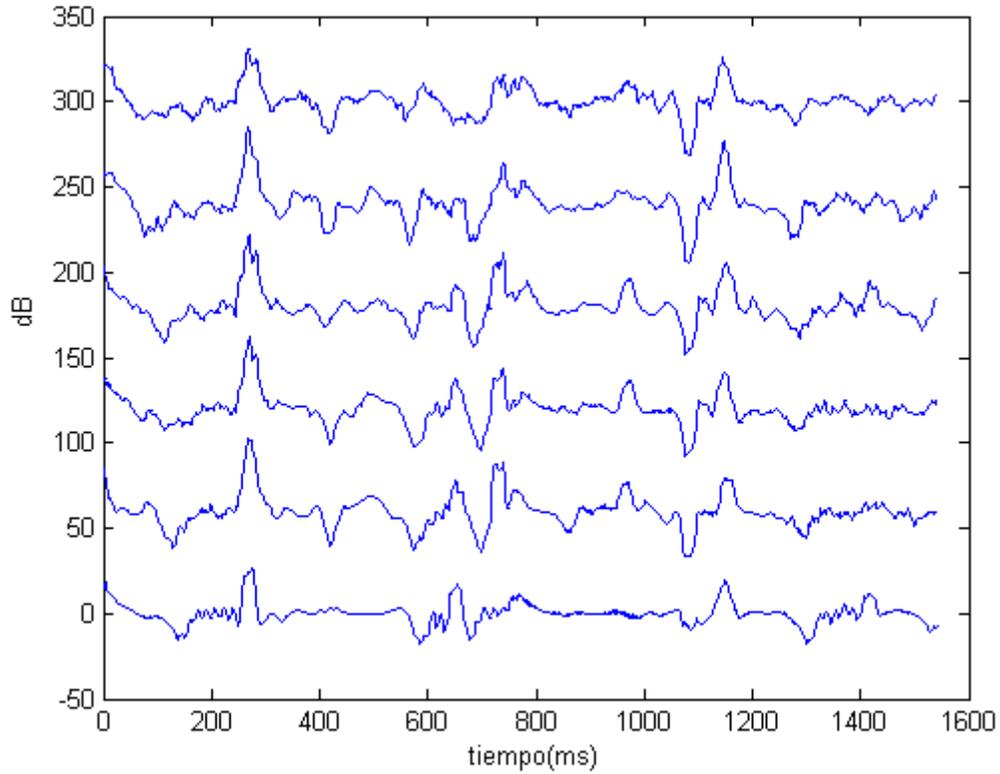


Figura A.2. Derivada de la energía en las seis bandas (*derivpf*).

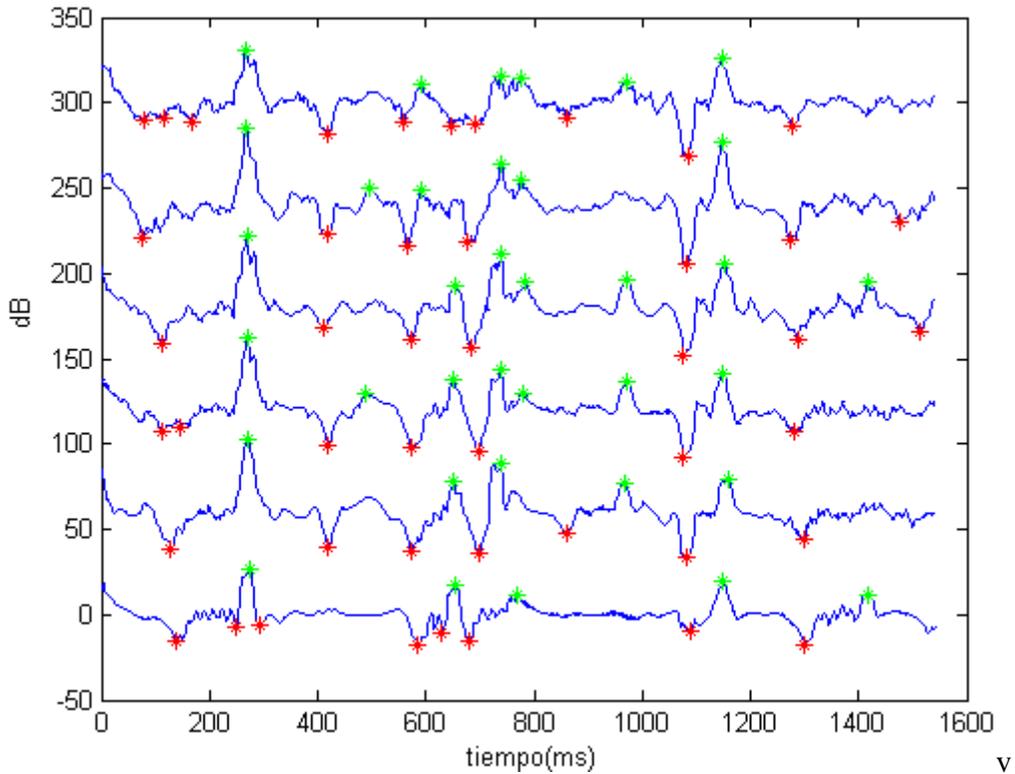


Figura A.3. Picos encontrados (*umbralpf*).

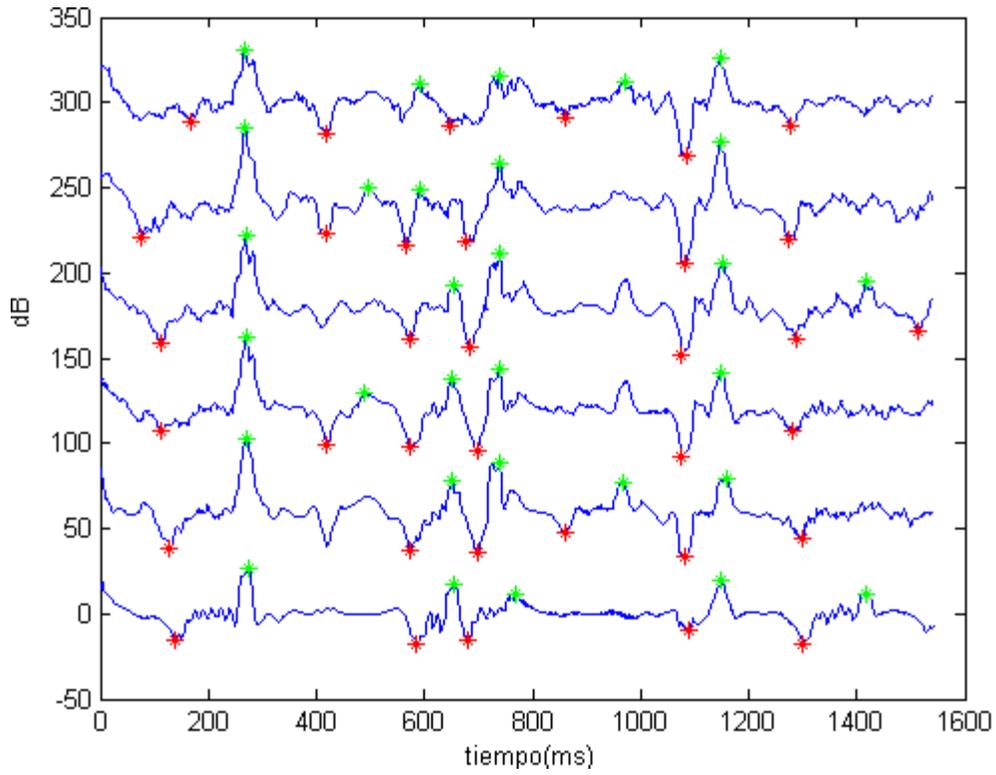


Figura A.4. Secuencia de picos (*secuenciapf*).

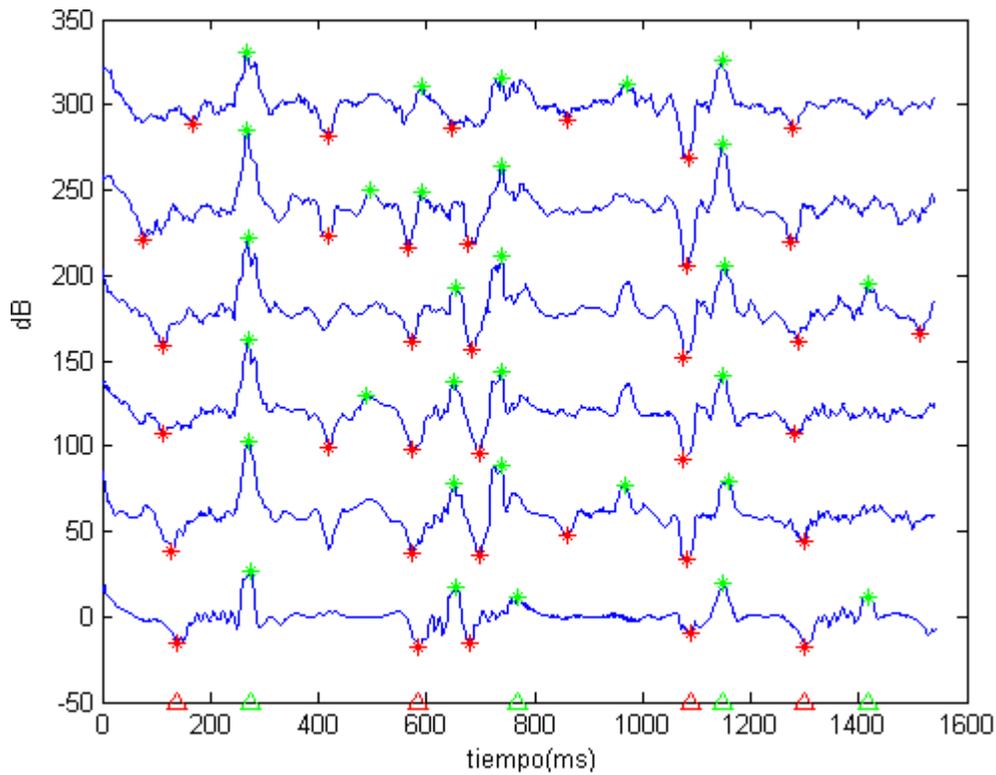


Figura A.5. Detección de marcas g (*glottispf*).

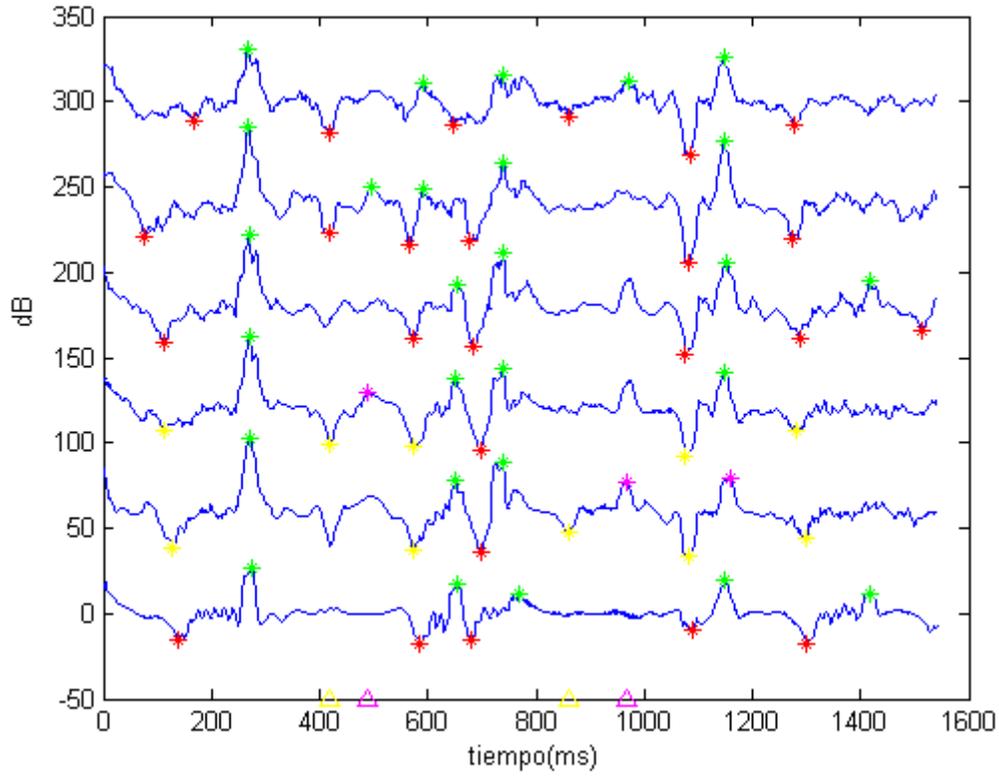


Figura A.6. Detección de marcas **s** (*sonorantpf*).

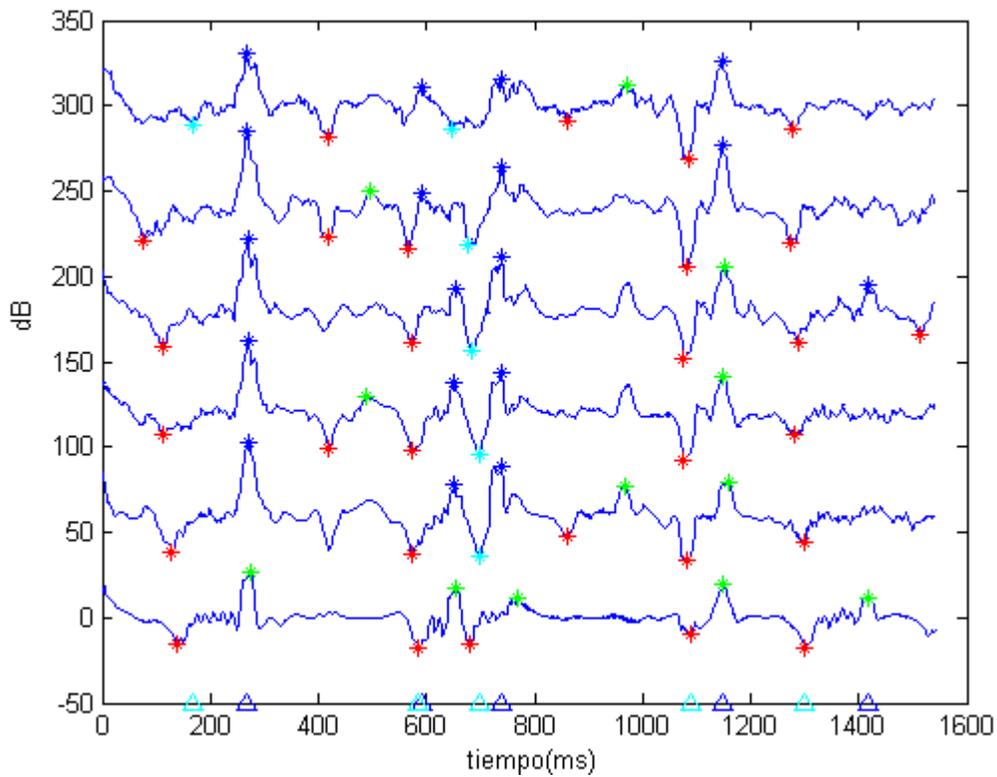


Figura A.7. Detección de marcas **b** (*burstpf*).

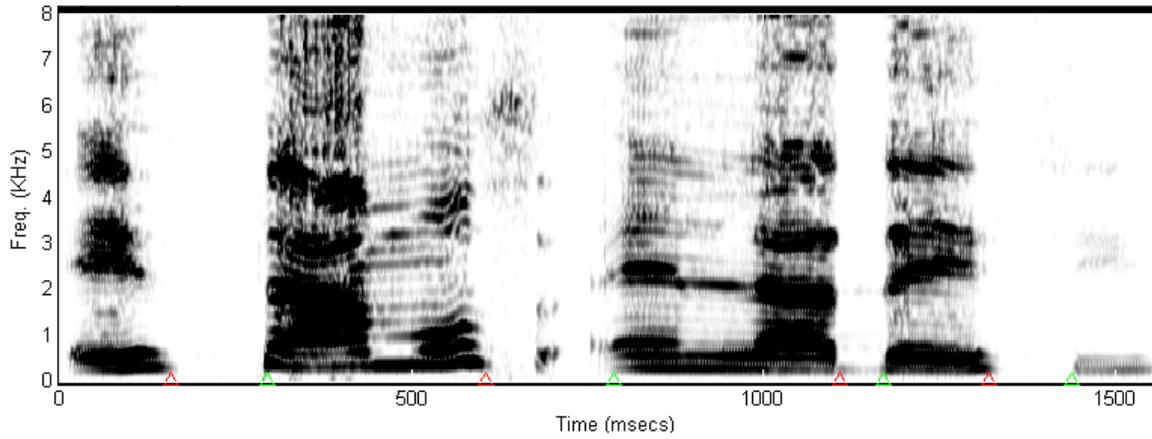


Figura A.8. Espectrograma y marcas **g** que se obtienen con la función *markspegpf*.

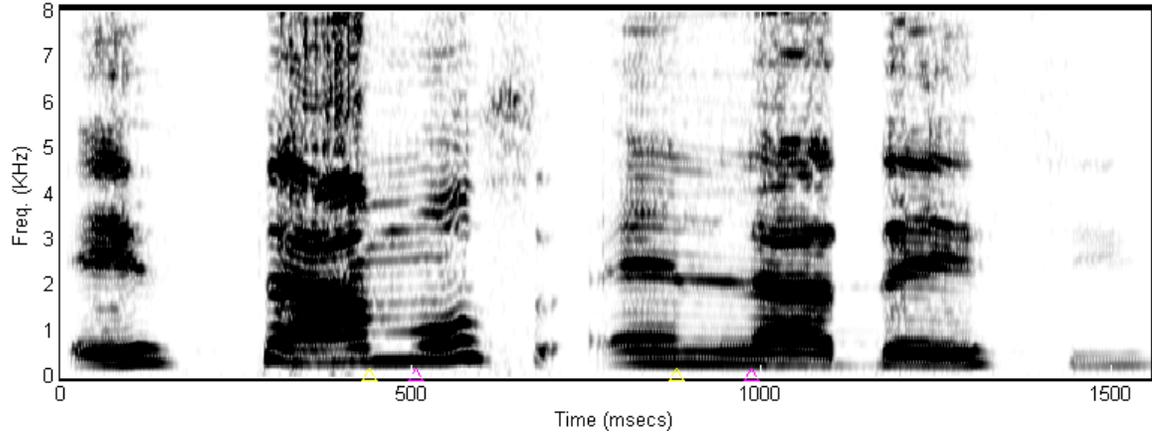


Figura A.9. Espectrograma y marcas **s** que se obtienen con la función *markspegpf*.

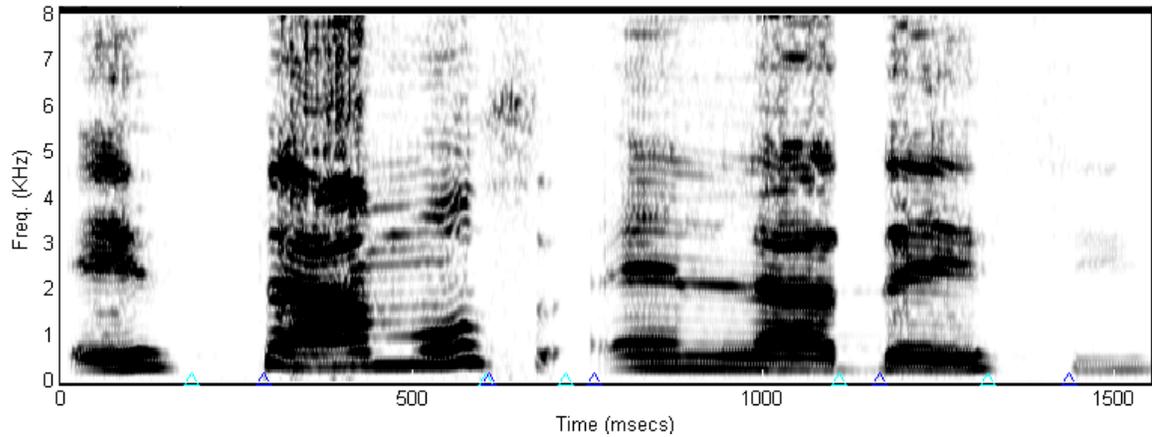


Figura A.10. Espectrograma y marcas **b** que se obtienen con la función *markspegpf*.

Anexo B Listado del programa en MATLAB versión 6.5

Las funciones *enbandapf*, *energiapf*, *derivpf* y *umbralpf* están recogidas en el fichero **edpf.m** y se ejecutan en el siguiente orden.

Enbandapf

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Define las seis bandas de frecuencia %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

nv = floor( 0.006*fs ); % Ventana de 6ms

d=floor( 0.001 *fs ); % Incremento de trama

mimatriz=enframe(x,hanning(nv),d); %Cada fila es una trama

espe=abs(fft(mimatriz,512,2));% Transformada rápida de Fourier con 512 puntos

paso=fs/512;% valor del paso

v1=floor(400/paso);% cantidad de pasos para cada frecuencia según las bandas

v2=floor(800/paso);

v3=floor(1500/paso);

v4=floor(1200/paso);

v5=floor(2000/paso);

v6=floor(3500/paso);

v7=floor(5000/paso);

v8=floor(8000/paso);

banda1=espe(:,1:12);% números de columnas tomadas para definir las bandas.

banda2=espe(:,26:48);

banda3=espe(:,38:64);

banda4=espe(:,64:112);

banda5=espe(:,112:160);
```

```
banda6=espe(:,160:256);
```

Energiapf

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Devuelve las variaciones de energía en las seis bandas calculada como la magnitud
% cuadrada de un intervalo de 10ms cada 1ms y tomando el máximo valor en cada paso %
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Banda 1 %
```

```
j=[];
```

```
E1=[];
```

```
for i=[1:size(banda1,1)-10]; %calculo de la energía.
```

```
    ventana20 = [i:i+9]; %suavizamiento de 10ms para el paso fino.
```

```
    bandam= (banda1(i:i+9,:));
```

```
    p=bandam.^2;
```

```
    meanban=mean(p,1);
```

```
    mx=max(meanban);
```

```
    e=mx;
```

```
    E1=[E1 e];
```

```
end;
```

```
figure;plot(10.*log10(E1));
```

Derivpf

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Devuelve la derivada de la energía para pasos de 26ms cada 1ms%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Banda 1 %
```

```
R1=[]; % Calculo de la derivada
```

```

E11=10.*log10(E1);
for i=[1:size(E11,2)-26]; % 26 ms es el paso fino
    resta1=E11(1,i+25)-E11(1,i); % primera derivada banda 1
    R1=[R1 resta1];
end;
figure;plot(R1);

```

Umbralpf

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Localiza los picos positivos y negativos presentes en el resultado anterior que son
% mayores y menores que 6 y -6 dB (para las bandas 2, 3, 4, 5 y 6 el umbral es 9 y -9dB)
% respectivamente, devolviendo la amplitud y localización del mayor de los positivos y
% menor de los negativos en cada intervalo de 26ms %

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Banda 1%

```

```

Rp1=R1;                                Mp1(noop1)=0;
nop1=find(Rp1<0);                        [kp1,vp1]=findpeaks(Mp1);
Rp1(nop1)=0;                             Z1=[];
[k1,v1]=findpeaks(Rp1);                  for i=[1:size(vp1,1)];
Mp1=[];                                   zp1=find(v1==vp1(i));
for i=[1:26:size(Rp1,2)-26];              Z1=[Z1 zp1];
    venp1=Rp1(1,i:i+25);                  end;
    mp1=max(venp1);                        Rn1=R1;
    Mp1=[Mp1 mp1];                         non1=find(Rn1>0);
end;                                        Rn1(non1)=0;
noop1=find(Mp1<6);                         Mn1=[];

```



```

% Banda 1 %
% Negativos %
M=[];
N=[];
for i=[1:length(L1)-1];
    gg=find(Ln1<L1(i+1));
    yy=Ln1(gg);
    gg1=find(yy>L1(i));
    if length(gg1)~=0;
        yy1=yy(gg1);
        G=[];
        for i=[1:length(gg1)];
            busca=find(Ln1==yy1(i));
            G=[G busca];
        end;
        rr=vn1(G);
        [xmin]=min(rr);
        rr1=find(vn1==xmin);
        N=[N rr1];
    end;
end;
if Ln1(1)<L1(1);
    uu=find(Ln1<L1(1));
    uu1=vn1(uu);
    [xmin]=min(uu1);
    uu2=find(vn1==xmin);
    N=[N uu2];
end;
ce=find(Ln1>L1(length(L1)));
if length(ce)~=0;
    ce1=vn1(ce);
    [xmin]=min(ce1);
    ce2=find(vn1==xmin);
    N=[N ce2];
end;
me11=Ln1(1,N);
mi11=vn1(N,1);
figure;plot(R1);
hold on;plot(me11,mi11,'r*');
% Positivos %
M=[];
N=[];
for i=[1:length(Ln1)-1];
    gg=find(L1<Ln1(i+1));
    yy=L1(gg);
    gg1=find(yy>Ln1(i));

```

```

if length(gg1)~=0;
    yy1=yy(gg1);
    G=[];
    for i=[1:length(gg1)];
        busca=find(L1==yy1(i));
        G=[G busca];
    end;
    rr=vp1(G);
    [xmax]=max(rr);
    rr1=find(vp1==xmax);
    M=[M rr1];
end;
end;
if L1(1)<Ln1(1);
    uu=find(L1<Ln1(1));
    uu1=vp1(uu);
    [xmax]=max(uu1);
    uu2=find(vp1==xmax);
    M=[M uu2];
end;
ce=find(L1>Ln1(length(Ln1)));
if length(ce)~=0;
    ce1=vp1(ce);
    [xmax]=max(ce1);
    ce2=find(vp1==xmax);
    M=[M ce2];
end;
me1=L1(1,M);
mi1=vp1(M,1);
hold on;plot(me1,mi1,'g*');
% Banda 2 %
n2=[];
for i=[1:length(me2)];
    xx=me2(1,i);
    n2=[n2 xx];
end;
end;
for i=[1:length(me12)];
    xx=me12(1,i);
    n2=[n2 xx];
end;
ss=sort(me1);
ss1=sort(me11);
SS=[];
SS1=[];
if ss1(1)<ss(1);
    SS1=[SS1 ss1(1)];
    if length(ss1)>length(ss);
        for i=[1:length(ss)];

```

<pre> cw=ss(i); cw1=ss1(i+1); cw2=cw1-cw; if cw2>=30; SS=[SS cw]; SS1=[SS1 cw1]; end; SS1=[SS1 cw1]; end; end; end; if length(ss1)<length(ss); for i=[1:length(ss)-1]; cw=ss(i); cw1=ss1(i+1); cw2=cw1-cw; if cw2>=30; SS=[SS cw]; SS1=[SS1 cw1]; end; end; end; if length(ss1)==length(ss); for i=[1:length(ss1)-1]; cw=ss(i); cw1=ss1(i+1); </pre>	<pre> cw2=cw1-cw; if cw2>=30; SS=[SS cw]; SS1=[SS1 cw1]; end; end; end; if ss1(1)>ss(1); if length(ss1)>length(ss); for i=[1:length(ss)]; cw=ss(i); cw1=ss1(i); cw2=cw1-cw; if cw2>=30; SS=[SS cw]; SS1=[SS1 cw1]; end; end; end; end; if length(ss1)<length(ss); for i=[1:length(ss1)]; cw=ss(i); cw1=ss1(i); </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

```

cw2=cw1-cw;                SS=[SS cw];
if cw2>=30;                SS1=[SS1 cw1];
    SS=[SS cw];            end;
    SS1=[SS1 cw1];        end;
end;                        end;
end;                        end;
end;                        if ss(length(ss))>SS1(length(SS1));
if length(ss1)==length(ss);    SS=[SS ss(length(ss))];
for i=[1:length(ss1)];        end;
    cw=ss(i);                ss=SS;
    cw1=ss1(i);              ss1=SS1;
    cw2=cw1-cw;
    if cw2>=30;

```

Glottispf

```

%%%%%%%%%%
% Esta función se efectúa después de las mencionadas anteriormente %
% Devuelve los valores de las marcas g positivas y negativas en los vectores glotp y glotn
% respectivamente %
%%%%%%%%%%
glotn=[];
glotp=[];
qg=[];
qg=zeros(1,length(ss1));
hold on;plot(ss1,qg-50,'r^');
glotn=[glotn ss1];

```

```

qg1=[];
qg1=zeros(1,length(ss));
hold on;plot(ss,qg1-50,'g^');
glotp=[glotp ss];

```

Sonorantpf

%%%%%%%%%%
 % Devuelve los valores de las marcas s positivas y negativas en los vectores sonop y sonon respectivamente. Estos valores se localizan en intervalos de segmentos sonoros y están delimitados por un pico negativo y uno positivo medidos en las bandas 2 y 3 %

%%%%%%%%%%

```

sonon=[];
sonop=[];
if ss(1)>ss1(1);
    z1n=me12(zn);
mz=[];
    z2n=find(z1n>u);
mzn=[];
    z3n=z1n(z2n);
o=ss1(1);
    mzn=[mzn z3n];
u=0;
    nzn=[];
% Banda 2 %
    for i=[1:length(mz)];
z=find(me2<o);
        z4=find(me2==mz(i));
z1=me2(z);
        nz=[nz z4];
z2=find(z1>u);
    end;
z3=z1(z2);
    mz1=mi2(nz);
mz=[mz z3];
    hold on; plot(mz,mz1+60,'m*');
nzn=[];
    for i=[1:length(mzn)];
zn=find(me12<o);
        z4n=find(me12==mzn(i));

```

```

    nzn=[nzn z4n];
end;
mz1n=mi12(nzn);
hold on;plot(mzn,mz1n+60,'y*');

% Banda 3 %
mz2=[];
z=find(me3<o);
z1=me3(z);
z2=find(z1>u);
z3=z1(z2);
mz2=[mz2 z3];
nz2=[];
for i=[1:length(mz2)];
    z4=find(me3==mz2(i));
    nz2=[nz2 z4];
end;
mz12=mi3(nz2);
hold on;plot(mz2,mz12+120,'m*');
mzn2=[];
zn=find(me13<o);
z1n=me13(zn);
z2n=find(z1n>u);
z3n=z1n(z2n);

    mzn2=[mzn2 z3n];
    nzn2=[];
    for i=[1:length(mzn2)];
        z4n=find(me13==mzn2(i));
        nzn2=[nzn2 z4n];
    end;
    mz1n2=mi13(nzn2);
    hold on;plot(mzn2,mz1n2+120,'y*');
    % posiciones positivos %
    nn=[];
    for i=[1:length(mz)];
        xx=mz(1,i);
        nn=[nn xx];
    end;
    for i=[1:length(mz2)];
        xx=mz2(i);
        nn=[nn xx];
    end;
    % posiciones negativos %
    nn1=[];
    for i=[1:length(mzn)];
        xx=mzn(i);
        nn1=[nn1 xx];
    end;

```

```

end;                                ww=[ww min(bb)];
for i=[1:length(mzn2)];              ww1=[ww1 ee];
    xx=mzn2(1,i);                    end;
    nn1=[nn1 xx];                    end;
end;                                  end;
if length(nn1)~=0 & length(nn)~=0;   q=[];
    aa=find(nn<min(nn1));            q1=[];
    gg=find(nn<min(nn1));            q=zeros(length(ww));
    cc=1;                             q1=zeros(length(ww1));
    ww=[];                             hold on;plot(ww,q-50,'m^');
    ww1=[];                             hold on;plot(ww1,q1-50,'y^');
    ww2=[];                             sonon=[sonon ww1];
    ww4=[];                             sonop=[sonop ww];
    yu=[];                               end;
    if length(aa)~=0;                 if length(gg)==0;
        while length(aa)~=0& length(cc)~=0;    aa1=find(nn1<min(nn));
            bb=nn(aa);                          while length(aa1)~=0 &
            cc=find(nn1>min(bb));                length(cc)~=0;
            if length(cc)~=0;                    bb=nn1(aa1);
                dd=nn1(cc);                      cc=find(nn>min(bb));
                ee=min(dd);                      if length(cc)~=0;
                aa=find(nn>ee);                  dd=nn(cc);
            if length(aa)~=0;                    ee=min(dd);
                bb=nn(aa);                      ww4=[ww4 ee];

```

```

ww2=[ww2 min(bb)];
aa1=find(nn1>ee);
end;
end;
q=[];
q1=[];
q=zeros(length(ww2));
q1=zeros(length(ww4));
hold on;plot(ww2,q-50,'y^');
hold on;plot(ww4,q1-50,'m^');
sonon=[sonon ww2];
sonop=[sonop ww4];
end;
end;
end;
val=find(ss1>ss(1));
vals=ss1(val);
ss1=vals;
for i=[1:length(ss)];
mz=[];
mzn=[];
o=ss1(i);
u=ss(i);
z=find(me2<o);
z1=me2(z);
z2=find(z1>u);
z3=z1(z2);
mz=[mz z3];
nz=[];
zn=find(me12<o);
z1n=me12(zn);
z2n=find(z1n>u);
z3n=z1n(z2n);
mzn=[mzn z3n];
nzn=[];
for i=[1:length(mz)];
z4=find(me2==mz(i));
nz=[nz z4];
end;
mz1=mi2(nz);
hold on; plot(mz,mz1+60,'m*');
for i=[1:length(mzn)];
z4n=find(me12==mzn(i));
nzn=[nzn z4n];
end;
mz1n=mi12(nzn);
hold on;plot(mzn,mz1n+60,'y*');
mz2=[];

```

```

z=find(me3<o);
z1=me3(z);
z2=find(z1>u);
z3=z1(z2);
mz2=[mz2 z3];
nz2=[];
for i=[1:length(mz2)];
    z4=find(me3==mz2(i));
    nz2=[nz2 z4];
end;
mz12=mi3(nz2);
hold on;plot(mz2,mz12+120,'m*');
mzn2=[];
zn=find(me13<o);
z1n=me13(zn);
z2n=find(z1n>u);
z3n=z1n(z2n);
mzn2=[mzn2 z3n];
nzn2=[];
for i=[1:length(mzn2)];
    z4n=find(me13==mzn2(i));
    nzn2=[nzn2 z4n];
end;
mz1n2=mi13(nzn2);

hold on;plot(mzn2,mz1n2+120,'y*')
nn=[];
for i=[1:length(mz)];
    xx=mz(1,i);
    nn=[nn xx];
end;
for i=[1:length(mz2)];
    xx=mz2(i);
    nn=[nn xx];
end;
nn1=[];
for i=[1:length(mzn)];
    xx=mzn(i);
    nn1=[nn1 xx];
end;
for i=[1:length(mzn2)];
    xx=mzn2(1,i);
    nn1=[nn1 xx];
end;
if length(nn1)~=0 & length(nn)~=0;
    aa=find(nn<min(nn1));
    gg=find(nn<min(nn1));
    cc=1;
    ww=[];

```

```

ww1=[];
ww2=[];
ww4=[];
yu=[];
    if length(aa)~=0;
while length(aa)~=0 & length(cc)~=0;
    bb=nn(aa);
    cc=find(nn1>min(bb));
    if length(cc)~=0;
        dd=nn1(cc);
        ee=min(dd);
        aa=find(nn>ee);
        if length(aa)~=0;
            bb=nn(aa);
            ww=[ww min(bb)];
            ww1=[ww1 ee];
        end;
    end;
end;
end;
q=[];
q1=[];
q=zeros(length(ww));
q1=zeros(length(ww1));
hold on;plot(ww,q-50,'m^');

hold on;plot(ww1,q1-50,'y^');
sonon=[sonon ww1];
sonop=[sonop ww];
end;
if length(gg)==0;
    aa1=find(nn1<min(nn));
    while length(aa1)~=0& length(cc)~=0;
        bb=nn1(aa1);
        cc=find(nn>min(bb));
        if length(cc)~=0;
            dd=nn(cc);
            ee=min(dd);
            ww4=[ww4 ee];
            ww2=[ww2 min(bb)];
            aa1=find(nn1>ee);
        end;
    end;
q=[];
q1=[];
q=zeros(length(ww2));
q1=zeros(length(ww4));
hold on;plot(ww2,q-50,'y^');
hold on;plot(ww4,q1-50,'m^');
sonon=[sonon ww2];

```

```
sonop=[sonop ww4];
end;
end;
end;
```

Burstpf

```
%%%%%%%%%%
% Devuelve los valores de las marcas b positivas y negativas en los vectores sorp y sorn
% respectivamente. Estos valores se localizan en intervalos de segmentos sordos y están
% delimitados por un pico negativo y uno positivo medidos desde la banda 2 a la 6 %
%%%%%%%%%%
% Se buscan los picos en las bandas 2, 3, 4, 5 y 6 que se encuentran entre una marca g
% negativo a la izquierda y una marca g positive a la derecha %
```

```
wy=find(ss1>ss1(1));          u=ss(i+1);
wy1=ss1(wy);                 % Banda 2 (se repite par alas demás
wp=ss(length(ss));          bandas) %
wn=find(wy1<wp);            z=find(me2>o);
wn1=wy1(wn);                z1=me2(1,z);
wy2=find(wy1<max(wn1));     z2=find(z1<u);
if length(wy2)~=0;          z3=z1(1,z2);
    for i=[1:length(wy2)];   mz=[mz z3];
        mz=[];              nz=[];
mzn=[];                      zn=find(me12>o);
o=ss1(i+1);                  z1n=me12(1,zn);
```

```

z2n=find(z1n<u);
z3n=z1n(1,z2n);
mzn=[mzn z3n];
nzn=[];
for i=[1:length(mz)];
    z4=find(me2==mz(i));
    nz=[nz z4];
end;
mz1=mi2(nz,1);
hold on;plot(mz,mz1+60,'b*');
for i=[1:length(mzn)];
    z4n=find(me12==mzn(i));
    nzn=[nzn z4n];
end;
mz1n=mi12(nzn,1);
hold on;plot(mzn,mz1n+60,'c*');
% positivos %
mn=[];
for i=[1:length(mz)];
    xx=mz(1,i);
    nn=[nn xx];
end;
% negativos %
nn1=[];
for i=[1:length(mzn)];
    xx=mzn(1,i);
    nn1=[nn1 xx];
end;
% Con los valores de los picos en cada
una de las bandas se buscan las marcas.
% u es el valor de inicio de sonoridad.%
% o es el valor de fin de sonoridad. %
my=[];
if length(nn1)==0 & length(nn)==0;
    tt1=find(n2<u);
    tt2=n2(tt1);
    maxo=max(tt2);
    hh=find(me12==maxo);
    if length(hh)~=0;
        my=[my hh];
    end;
    tt3=find(n3<u);
    tt2=n3(tt3);
    maxo=max(tt2);
    hh=find(me13==maxo);
    if length(hh)~=0;
        my=[my hh];
    end;
end;

```

```

tt4=find(n4<u);
hold on;plot(u,q1-50,'b^');
tt2=n4(tt4);
hold on;plot(o,q1-50,'c^');
maxo=max(tt2);
sorn=[sorn o];
hh=find(me14==maxo);
sorp=[sorp u];
if length(hh)~=0;
end;
my=[my hh];
end;
if length(nn1)==0 & length(nn)~=0;
mu=[];
tt6=find(n5<u);
tt1=find(n2<=o);
tt2=n2(tt1);
maxo=max(tt2);
hh=find(me15==maxo);
maxo=max(tt2);
if length(hh)~=0;
hh=find(me12==maxo);
if length(hh)~=0;
mu=[mu hh];
end;
tt7=find(n6<u);
end;
tt2=n6(tt7);
tt3=find(n3<=o);
maxo=max(tt2);
tt2=n3(tt3);
hh=find(me16==maxo);
maxo=max(tt2);
if length(hh)~=0;
hh=find(me13==maxo);
if length(hh)~=0;
mu=[mu hh];
end;
if length(my)==5;
end;
q1=[];
tt4=find(n4<=o);
q1=zeros(1);
tt2=n4(tt4);

```

```

maxo=max(tt2);
hh=find(me14==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
tt6=find(n5<=o);
tt2=n5(tt6);
maxo=max(tt2);
hh=find(me15==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
tt7=find(n6<=o);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
if length(mu)==5;
    q1=[];
    q1=zeros(1);
    fmin=min(nn);
    hold on;plot(fmin,q1-50,'b^');
end;
hold on;plot(o,q1-50,'c^');
sorn=[sorn o];
sorp=[sorp fmin];
end;
mu=[];
tt1=find(n2<u);
tt2=n2(tt1);
maxo=max(tt2);
hh=find(me2==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
tt3=find(n3<u);
tt2=n3(tt3);
maxo=max(tt2);
hh=find(me3==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
tt4=find(n4<u);
tt2=n4(tt4);
maxo=max(tt2);
hh=find(me4==maxo);
if length(hh)~=0;

```

```

    mu=[mu hh];
end;
tt6=find(n5<u);
tt2=n5(tt6);
maxo=max(tt2);
hh=find(me5==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
tt7=find(n6<u);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me6==maxo);
if length(hh)~=0;
    mu=[mu hh];
end;
if length(mu)==5;
    q2=[];
    q2=zeros(1);
    fmin=min(nn);
    hold on;plot(fmin,q1-50,'b^');
    hold on;plot(o,q1-50,'c^');
    sorn=[sorn o];
    sorp=[sorp fmin];
end;
end;
end;
if length(nn1)~=0 & length(nn)==0;
    mt=[];
    tt1=find(n2<o);
    tt2=n2(tt1);
    maxo=max(tt2);
    hh=find(me2==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
    tt3=find(n3<o);
    tt2=n3(tt3);
    maxo=max(tt2);
    hh=find(me3==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
    tt4=find(n4<o);
    tt2=n4(tt4);
    maxo=max(tt2);
    hh=find(me4==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
end;

```

```

end;                                mt=[];
tt6=find(n5<o);                      tt1=find(n2<u);
tt2=n5(tt6);                         tt2=n2(tt1);
maxo=max(tt2);                       maxo=max(tt2);
hh=find(me5==maxo);                  hh=find(me12==maxo);
if length(hh)~=0;                    if length(hh)~=0;
    mt=[mt hh];                       mt=[mt hh];
end;                                  end;
tt7=find(n6<o);                      tt3=find(n3<u);
tt2=n6(tt7);                         tt2=n3(tt3);
maxo=max(tt2);                       maxo=max(tt2);
hh=find(me6==maxo);                  hh=find(me13==maxo);
if length(hh)~=0;                    if length(hh)~=0;
    mt=[mt hh];                       mt=[mt hh];
end;                                  end;
if length(mt)==5;                    tt4=find(n4<u);
    q2=[];                             tt2=n4(tt4);
    q2=zeros(1);                       maxo=max(tt2);
    fmax=max(nn1);                     hh=find(me14==maxo);
    hold on;plot(fmax,q2-50,'c^');      if length(hh)~=0;
    hold on;plot(u,q2-50,'b^');         mt=[mt hh];
    sorn=[sorn fmax];                 end;
    sorp=[sorp u];                    tt6=find(n5<u);
end;                                  tt2=n5(tt6);

```

```

maxo=max(tt2);
hh=find(me15==maxo);
if length(hh)~=0;
    mt=[mt hh];
end;
tt7=find(n6<u);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    mt=[mt hh];
end;
if length(mt)==5;
    q2=[];
    q2=zeros(1);
    fmax=max(nn1);
    hold on;plot(fmax,q2-50,'c^')
    hold on;plot(u,q2-50,'b^');
    sorn=[sorn fmax];
    sorp=[sorp u];
end;
end;
my=[];
if length(nn1)~=0 & length(nn)~=0;
    tt1=find(n2<o);
    tt2=n2(tt1);
    maxo=max(tt2);
    hh=find(me12==maxo);
    if length(hh)~=0;
        my=[my hh];
    end;
    tt3=find(n3<o);
    tt2=n3(tt3);
    maxo=max(tt2);
    hh=find(me13==maxo);
    if length(hh)~=0;
        my=[my hh];
    end;
    tt4=find(n4<o);
    tt2=n4(tt4);
    maxo=max(tt2);
    hh=find(me14==maxo);
    if length(hh)~=0;
        my=[my hh];
    end;
    tt6=find(n5<o);
    tt2=n5(tt6);
    maxo=max(tt2);

```

```

hh=find(me15==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt7=find(n6<o);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    my=[my hh];
end;
if length(my)==5;
    q1=[];
    q1=zeros(1);
    hold on;plot(o,q1-50,'c^');
    hold on;plot(min(nn),q1-50,'b^');
    sorn=[sorn o];
    sorp=[sorp min(nn)];
end;
neg=find(nn1<u);
if length(neg)~=0;
    neg1=nn1(neg);
    ne=max(neg1);
    mt=[];
    tt1=find(n2<=ne);
    tt2=n2(tt1);
    maxo=max(tt2);
    hh=find(me12==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
    tt3=find(n3<=ne);
    tt2=n3(tt3);
    maxo=max(tt2);
    hh=find(me13==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
    tt4=find(n4<=ne);
    tt2=n4(tt4);
    maxo=max(tt2);
    hh=find(me14==maxo);
    if length(hh)~=0;
        mt=[mt hh];
    end;
    tt6=find(n5<=ne);
    tt2=n5(tt6);
    maxo=max(tt2);

```

```

hh=find(me15==maxo);
if length(hh)~=0;
    mt=[mt hh];
end;
tt7=find(n6<=ne);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    mt=[mt hh];
end;
if length(mt)==5;
    posi=find(nn>ne);
    if length(posi)~=0;
        posi1=nn(posi);
        q2=[];
        q2=zeros(1);
        hold on;plot(ne,q2-50,'c^');
        hold on;plot(min(posi1),q2-50,'b^');
        sorn=[sorn ne];
        sorp=[sorp min(posi1)];
    end;
if length(posi)==0;
    q2=[];

```

```

q2=zeros(1);
hold on;plot(ne,q2-50,'c^');
hold on;plot(u,q2-50,'b^');
sorn=[sorn ne];
sorp=[sorp u];
end;
end;
my=[];
tt1=find(n2<min(nn));
tt2=n2(tt1);
maxo=max(tt2);
hh=find(me12==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt3=find(n3<min(nn));
tt2=n3(tt3);
maxo=max(tt2);
hh=find(me13==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt4=find(n4<min(nn));
tt2=n4(tt4);

```

```

maxo=max(tt2);
hh=find(me14==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt6=find(n5<min(nn));
tt2=n5(tt6);
maxo=max(tt2);
hh=find(me15==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt7=find(n6<min(nn));
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    my=[my hh];
end;
if length(my)==5;
    cu=find(nn1<min(nn));
if length(cu)~=0;
    cu1=nn1(cu);
    q1=[];
    q1=zeros(1);
    hold on;plot(min(nn),q1-50,'b^');
    hold on;plot(max(cu1),q1-50,'c^');
    sorn=[sorn max(cu1)];
    sorp=[sorp min(nn)];
end;
end;
if length(cu)==0;
    q1=[];
    q1=zeros(1);
    hold on;plot(min(nn),q1-50,'b^');
    hold on;plot(o,q1-50,'c^');
    sorn=[sorn o];
    sorp=[sorp min(nn)];
end;
end;
my=[];
tt1=find(n2<u);
tt2=n2(tt1);
maxo=max(tt2);
hh=find(me12==maxo);
if length(hh)~=0;
    my=[my hh];
end;
end;
tt3=find(n3<u);

```

```

tt2=n3(tt3);
maxo=max(tt2);
hh=find(me13==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt4=find(n4<u);
tt2=n4(tt4);
maxo=max(tt2);
hh=find(me14==maxo);
if length(hh)~=0;
    my=[my hh];
end;
tt6=find(n5<u);
tt2=n5(tt6);
maxo=max(tt2);
hh=find(me15==maxo);
if length(hh)~=0;
    my=[my hh];
end;

tt7=find(n6<u);
tt2=n6(tt7);
maxo=max(tt2);
hh=find(me16==maxo);
if length(hh)~=0;
    my=[my hh];
end;
if length(my)==5;
    q1=[];
    q1=zeros(1);
    hold on;plot(u,q1-50,'b^');
    hold on;plot(max(nn1),q1-50,'c^');
    sorn=[sorn max(nn1)];
    sorp=[sorp u];
end;
end;
end;
end;

```

Markspecgpf

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

% Esta función se ejecuta cuando han sido efectuadas todas las funciones anteriores %

```

```

% Devuelve el espectrograma de la señal de entrada con los diferentes tipos de marcas y la
diferencia entre ellas %

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Marcas g %
figure;specgram(x,[],16000);colormap(gra
ay);
qg=[];
qg=zeros(1,length(glotn));
hold on;plot(glotn/1000,qg,'r^');
qg1=[];
qg1=zeros(1,length(glotp));
hold on;plot(glotp/1000,qg1,'g^');
difg=[];
for i=[1:length(glotn)];
    rc=glotn(i);
    rc1=glotp(i);
    diferencia=rc-rc1;
    difg=[difg diferencia];
end;
% Marcas s %
figure;specgram(x,[],16000);colormap(gra
ay);
qg=[];
qg=zeros(1,length(sonon));
hold on;plot(sonon/1000,qg,'y^');
qg1=[];
qg1=zeros(1,length(sonop));
hold on;plot(sonop/1000,qg1,'m^');
difs=[];
for i=[1:length(sonon)];
    rc=sonon(i);
    rc1=sonop(i);
    diferencia=rc1-rc;
    difs=[difs diferencia];
end;
% Marcas b %
figure;specgram(x,[],16000);colormap(gra
ay);
qg=[];
qg=zeros(1,length(sorn));
hold on;plot(sorn/1000,qg,'c^');
qg1=[];
qg1=zeros(1,length(sorp));
hold on;plot(sorp/1000,qg1,'b^');
difsor=[];
for i=[1:length(sorn)];
    rc=sorn(i);
    rc1=sorp(i);
    diferencia=rc1-rc;
    difsor=[difsor diferencia];
end;

```