

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnología de la Información



TRABAJO DE DIPLOMA

Detector de consonantes plosivas.

Autor: Esteban E. Barrizonte Anaya

Tutor: Ing. Héctor A. Kairuz Hernández – Díaz

Santa Clara

2012

"Año 54 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

Detector de consonantes plosivas.

Autor: Esteban E. Barrizonte Anaya.

Tutor: Ing. Héctor A. Kairuz Hernández - Díaz

Reserva Científica, Centro de Estudios de Electrónica y Tecnologías de la Información, Facultad Ing. Eléctrica. UCLV.

E-mail: akairuz@uclv.edu.cu

Santa Clara

2012

"Año 54 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 Biomédica, 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

La fe mística no es un medio para llegar a la verdad, sino para oscurecerla y detenerla; no ayuda al hombre, sino que lo detiene; no le responde, sino que lo castiga; no le satisface, sino que lo irrita.

José Martí.

DEDICATORIA

A mi madre que es lo que más quiero en esta vida, a mi padre Vitalio por estar siempre a mi lado dándome los mejores consejos, a mi hermano Rodolfo que en paz descansa y que Dios lo tenga en la gloria, a mis amigos en general, a toda mi familia y en especial a mi abuelo.

AGRADECIMIENTOS

A mi tutor y amigo Héctor A. Kairuz por su dedicación y por toda su paciencia.

A toda mi familia y en especial a mi madre por apoyarme en todo momento, a mi padre Vitalio por confiar siempre en mí y a mi abuelo por darme tanto animo.

A todos mis colegas por compartir buenos momentos conmigo y en especial a mi mejor amigo y hermano ‘El Rafa’ alias ‘Pepe’.

A mis mejores amigas Aniladys, Diana, Lianet, Irina, Liliana y Zumito por hacerme pasar ratos agradables a su lado.

A todas esas personas que aportaron su granito de arena para que yo pudiera salir adelante.

En verdad muchas gracias a todos.

TAREA TÉCNICA

1. Revisión bibliográfica sobre los algoritmos de detección de plosivas.
2. Programación de un algoritmo para localizar las plosivas en la señal de habla.
3. Análisis del desempeño del algoritmo en una muestra representativa.
4. Confección del informe final.

Firma del Autor

Firma del Tutor

RESUMEN

La carencia de medidas objetivas en el ejercicio Diadocoquinesis impone la necesidad de crear las bases para aplicar un conjunto de medidas acústicas que se correlacionen con las evaluaciones subjetivas de los especialistas.

Por lo anterior, se lleva a cabo un análisis de varios tipos de algoritmos, validados en la literatura, en cuanto a sus características, para realizar la implementación de un algoritmo que permita la detección de consonantes plosivas.

Como aspecto novedoso, se efectúa de forma automática las localizaciones de los puntos en la señal del habla donde ocurren consonantes oclusivas. Por último, se ilustran las pruebas realizadas con el algoritmo modificado sobre una muestra donde se detectan las consonantes plosivas que ocurren durante el ejercicio Diadocoquinesis.

TABLA DE CONTENIDOS

PENSAMIENTO	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
TAREA TÉCNICA	iv
RESUMEN.....	v
TABLA DE CONTENIDOS	vi
INTRODUCCIÓN	1
Organización del informe.....	2
CAPÍTULO 1. Medidas relacionadas con las consonantes plosivas y la articulación.	3
1.1 Antecedentes.	3
1.2 Las consonantes plosivas.	4
1.2.1 Concepto de plosivas	4
1.2.2 Características Generales	4
1.2.3 Particularidades de las consonantes plosivas.	4
1.2.4 Características acusticas.....	5
1.3 Funciones del Tracto Vocal para la Articulación de Consonantes: (Sistema Supra Laríngeo).....	7
1.3.1 Función Velofaríngea.....	9
1.3.2 Coordinación de la actividad laríngea y supralaríngea.	9
1.3.3 Duración de la fase de explosión.	10
1.4 Ejercicio de Diadocoquinesis	10
1.5 Algoritmos de detección de plosivas.....	11
1.5.1 Detector G (glottis).	13

1.5.2	Detector B (burst).....	13
1.5.3	Detector c (oclusive)	14
1.6	Consideraciones finales.....	15
CAPÍTULO 2. Materiales y Métodos		16
2.1	Diseño del algoritmo	16
2.1.1	Detector de actividad Glotal	19
2.1.2	Detector de plosivas (b)	23
2.2	El experimento	25
2.3	Conclusiones del capítulo.....	25
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN		26
3.1	Consonante /ka/.....	26
3.2	Consonante /pa/.....	27
3.3	Consonante /ta/.....	28
3.4	Conclusiones del capítulo.....	29
CONCLUSIONES Y RECOMENDACIONES		30
	Conclusiones	30
	Recomendaciones.....	30
REFERENCIAS BIBLIOGRÁFICAS.....		32
Anexo I	Ejemplo de corrida para un paciente.	37

INTRODUCCIÓN

En el siglo XIX aparecen investigaciones donde expertos utilizan el habla como medio de diagnóstico. En la segunda mitad de dicho siglo los estudios subjetivos de la voz comenzaron a ser aceptados, destacándose en esta etapa los reconocidos trabajos de las Clínicas de Mayo, donde se localiza la lesión neurológica a partir de las manifestaciones en el habla y se caracterizan los distintos tipos de disartria.

En la segunda mitad del siglo XX se busca que el diagnóstico esté menos afectado por quien lo realiza (disminuir la componente subjetiva), es por ello que se inicia junto con el desarrollo del procesamiento digital de señales, la era de mediciones acústicas.

En el presente siglo las medidas se concentran en buscar una integración hacia una medida objetiva de la inteligibilidad, lo cual requiere de medidas de articulación, nasalidad y prosodia, que en su mayoría, se han calculado en segmentos cortos debido a la complejidad del análisis del ejercicio Diadocoquinesis.

El ejercicio Diadocoquinesis es empleado para de comunicación natural, empleado para realizar las medidas subjetivas y a él se deben acercar las medidas objetivas, con el propósito de un mejor análisis del fenómeno del habla de forma global.

Hipótesis.

Se puede crear una herramienta a partir de los métodos de procesamiento digital del habla, que permita la detección de consonantes plosivas.

Objetivo General.

- Diseñar un detector de consonantes plosivas.

Objetivos Específicos.

- Realizar un análisis crítico de las posibles vías para la detección de consonantes plosivas.
- Implementar y comprobar el correcto funcionamiento del algoritmo para la detección de consonantes plosivas.
- Realizar pruebas para la validación del detector.

Organización del informe

El Capítulo 1 se dedica a la caracterización del problema a partir de un análisis de la literatura. Se comentan varios tipos de medidas para la detección de plosivas.

El Capítulo 2 explica el diseño metodológico de la investigación y se describe el algoritmo diseñado para la detección de plosivas.

En el Capítulo 3 se realiza la validación de la efectividad del algoritmo modificado mediante un análisis objetivo de los resultados con los datos adquiridos y las evaluaciones subjetivas, realizadas por expertos, en la muestra seleccionada.

CAPÍTULO 1. Medidas relacionadas con las consonantes plosivas y la articulación.

Resumen: En este capítulo se hace un análisis de los diferentes aspectos de las consonantes plosivas, sus características, la barra de explosión y VOT (Voice Onset Time por sus siglas en inglés) en personas normales. Se menciona las características del tracto vocal en la articulación de consonantes. También se comenta el principio de funcionamiento de los algoritmos de detección de plosivas.

1.1 Antecedentes.

Las plosivas han sido objeto de varios análisis acústicos en algunas de sus variedades ([1];[2];[3];[4];[5];[6];[7];[8];[9];[10];[11]); muchas de estas investigaciones se han orientado hacia la determinación de los parámetros que permiten diferenciar las oclusivas de acuerdo con el lugar de articulación (bilabial, dental, velar) o por las categorías sordas/sonoras.

El primer estudio encontrado es el trabajo de Borzone *et al* [2] en 1980, que hace una caracterización general de las plosivas en su variedad argentina en un corpus leído. Los autores trabajan duración del silencio, intensidad del ruido de explosión, VOT y el punto de partida de las transiciones formánticas en F2 y F3.

Posteriormente, Poch [10] en 1984 trabaja las plosivas en el español peninsular también en un corpus de habla controlada; analiza el parámetro duración de la oclusión, de la explosión y del VOT y las transiciones formánticas en el contexto intervocálico. Por otra parte, [3] en 1986 estudia específicamente el VOT de las oclusivas sordas y sonoras en la variedad peninsular, en un corpus de lectura de palabras, y compara sus resultados con los obtenidos para el inglés.

1.2 Las consonantes plosivas.

1.2.1 Concepto de plosivas

Termino utilizado en lugar de oclusiva particularmente en consonantes donde hay una liberación de aire luego de un bloqueo del tracto vocal.

1.2.2 Características Generales

Primero, cabe destacar que las consonantes oclusivas se perciben fácilmente en un espectrograma, dado que presentan una zona de silencio (de unos 90 ms. para las sordas y de unos 60 ms. para las sonoras) y una barra de explosión en el momento de la relajación. Como se muestra en la Figura 1 .

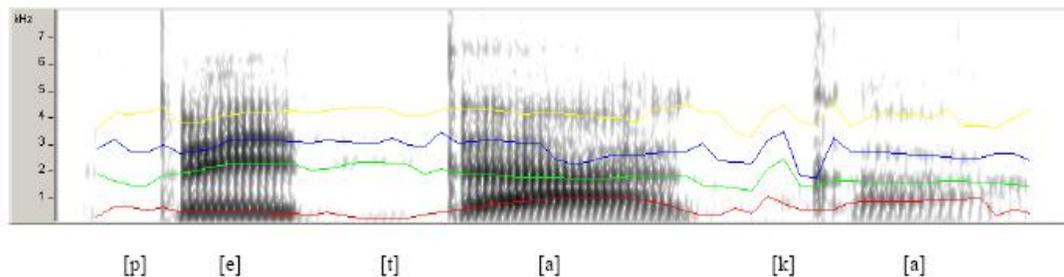


Figura 1 Espectrograma de banda ancha de las consonantes plosivas.

Las consonantes oclusivas sordas se caracterizan, además, por el hecho de que las cuerdas vocales no vibran durante la producción de las mismas. A su vez, otro rasgo fundamental de las consonantes oclusivas es que todas tienen barra de explosión y en el espectrograma se observa que carecen de barra de sonoridad.

1.2.3 Particularidades de las consonantes plosivas.

1.2.3.1 Características articulatorias

- Zona de articulación

[p] consonante bilabial: Se articula juntando o aproximando los dos labios.

[t]: consonante dental: Se trata de un sonido articulado en los dientes, en este caso, situando el ápice de la lengua entre los incisivos superiores e inferiores.

[k] consonante velar: Se articula en la zona del velo del paladar.

- Modo de articulación

En ellas los órganos crean un obstáculo a la salida del aire que dura milésimas de segundo, y suele deshacerse en una pequeña explosión.

1.2.4 Características acústicas.

Las consonantes oclusivas muestran en el espectrograma un espacio en blanco correspondiente al tiempo en que los órganos articulatorios se unen e impiden la salida al exterior del flujo del aire procedente de los pulmones. Si el sonido oclusivo es sonoro la vibración de las cuerdas vocales se refleja en el espectrograma con la presencia de una barra de sonoridad, una franja gruesa similar a un formante de frecuencia baja y ubicada inmediatamente encima del eje de abscisas.

En algunos contextos la explosión que acompaña a las oclusivas no se produce. Pero en caso de existir, figura en el espectrograma como una base perpendicular, llamada ‘barra de explosión’, localizada al final del sonido oclusivo y antes del sonido vocálico siguiente. Así, el punto de mayor intensidad de dicha barra puede estar situado a diferente altura con respecto al eje de ordenadas, lo que supone un indicio valioso para la identificación de la consonante de la que se trate.

Atendiendo a la oposición de los rasgos fónicos distintivos, también se observan diferencias notables entre las oclusivas sordas. De este modo, en la consonante velar [k], al ser densa, predomina más la zona central en el espectrograma que las regiones no centrales. En el caso de la labial [p] y la dental [t], ocurre lo contrario, destacando más las áreas periféricas. En lo que atañe a la oposición grave/agudo, en el sonido labial [p] y en el velar [k], la energía aparece distribuida en la zona baja del espectrograma, al tratarse de consonantes graves. En cambio, la consonante dental [t] es aguda, por lo que se da la situación contraria. Finalmente, en cuanto a la oposición estridente/mate, los sonidos estridentes se caracterizan por tener ondas sonoras de forma irregular, representadas por áreas de frecuencia distribuidas desordenadamente en el espectrograma. Las oclusivas se incluyen dentro de los sonidos mates, y por consiguiente, al poseer ondas más regulares, las zonas de frecuencia forman estrías verticales u horizontales.

Particularmente, comparando las tres oclusivas, se observan en primer lugar, diferencias respecto a la barra de explosión. La duración de la explosión permite determinar fácilmente el punto de articulación, de forma que en [k] la barra se ve más larga y visible que en [t], y a su vez en [t] se halla más marcada que en [p]. Así, en la bilabial apenas se aprecia la explosión, llegando a estar incluso ausente. Por otra parte, también conviene mirar la frecuencia en donde se sitúa la máxima intensidad de la explosión: en [p] los valores no llegan a los 1000 Hz; en [t] se localizan por encima de los 2500 Hz y en [k] oscilan entre 1000 y 3000 Hz dependiendo de la vocal siguiente. Las transiciones de las vocales vecinas son esclarecedoras: en la labial F2 y F3 son ascendentes; en la dental F2 es descendente y F3 ligeramente ascendente; y en la velar F2 es descendente, pero con una pendiente más pronunciada que en la dental, y F3 ligeramente ascendente. Figura 2 Espectrograma de banda ancha de la consonantes plosivas [pa] [ta].Figura 2 y Figura 3.

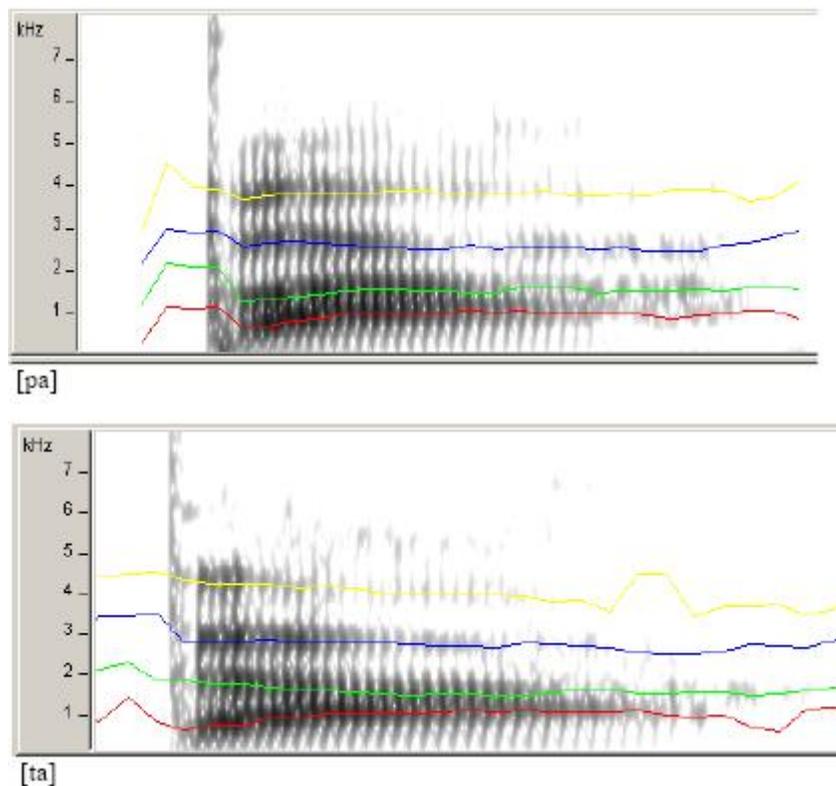


Figura 2 Espectrograma de banda ancha de la consonantes plosivas [pa] [ta].

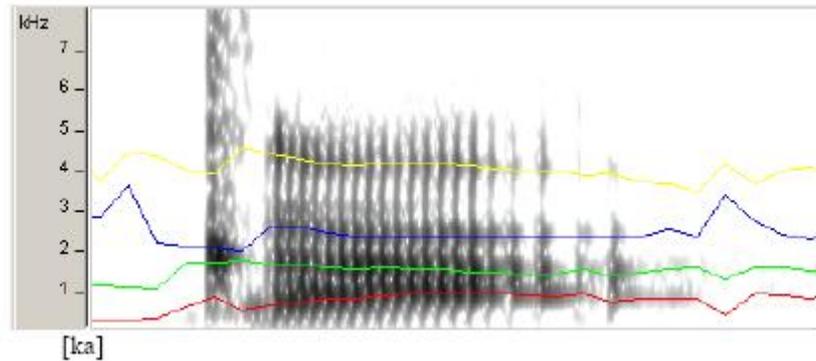


Figura 3 Espectrograma de banda ancha de la consonantes plosivas [ka].

1.3 Funciones del Tracto Vocal para la Articulación de Consonantes: (Sistema Supra Laríngeo).

Las Medidas sugeridas en [12] fueron: Espectro de oclusivas sordas o ruido de fricación.

Transiciones de formantes para consonante vocal (CV) y vocal consonante (VC).

Como las consonantes son una clase compleja de sonido, no hay una medida que diferencie cada clase; una manera útil de distinguirlas es en sonoras y sordas, basándose en el grado de constricción del tracto vocal.

Una consonante sorda está hecha por un cierre completo o un cierre estrecho en la cavidad oral de manera tal que el flujo de aire es detenido o se produce ruido fricativo.

Una consonante sonora es producida por un tracto vocal relativamente abierto por lo tanto ocurre resonancia en la cavidad. Las sordas incluyen fricativas, africadas y oclusivas. Las sonoras incluyen nasales, líquidas y semivocales. Las sonoras pueden describirse por patrones de formantes y antiformantes en estado estable y segmentos de transición. Los datos de estas son similares que para las vocales. Las sordas involucran en si un evento de fricación. Una explosión o ruido de transición para las oclusivas, un breve intervalo de ruido para las africadas y un largo intervalo de ruido para las fricativas.

Las mediciones en el dominio del tiempo están más correlacionadas que las espectrales, Aunque no hay medidas que representen en pocos términos el ruido espectral. Algunos trabajos han utilizado los momentos estadísticos. Estos son:

Primer momento (media) Da el centro de gravedad para el ruido y parece ser sensible a mala articulación de fricativas.

Segundo momento (desviación estándar) representa la distribución de energía alrededor de la media, una medida de los valores esparcidos alrededor de la media, su fórmula está dada por la Ecuación 1:

$$\text{Standard deviation} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Ecuación 1

Tercer momento (skewness) describe la simetría de la distribución. Su fórmula es la siguiente Ecuación 2:

$$\text{skewness} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n - 1} / s^3$$

Ecuación 2

La medida estadística skewness es cuantificada dividiendo el promedio elevado al cubo de las distancias por el cubo de la desviación estándar, valores mayores que cero indican skew positivas y menores que cero skew negativa.

Cuarto momento (kurtosis) está relacionado con la distribución de los picos de energía. Su fórmula es la siguiente Ecuación 3:

$$\text{kurtosis} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n - 1} / s^4$$

Ecuación 3

Esta ecuación es semejante a la de skewness excepto que las distancias de la media están elevadas a la cuarta potencia y el promedio de las distancias es dividida por la desviación estándar elevado a la cuarta potencia.

En uno de los pocos estudios donde los momentos espectrales fueron aplicados en la disartria, [13] se compararon los espectros fricativos del habla de personas con ALS (Amyotrophic Lateral Sclerosis) con el habla de personas neurológicamente sanas. La diferencia en frecuencia para el primer momento fue correlacionada con evaluaciones perceptuales de la precisión de la consonante. Se demostró que el primer momento (media espectral) es un índice útil de producción fricativa, particularmente cuando es combinado con la medida de la duración y la energía del segmento de ruido.

La kurtosis se correlacionó con problemas articulatorios en varios grupos disártricos en [14] y [15]. La precisión de la producción de las consonantes oclusivas puede ser determinada en parte por medidas de energía durante la fase oclusiva [16]. Algunos pacientes disártricos tienden a producir energía durante el silencio.

Esta energía tiene regularmente dos formas: ruido turbulento (espirantización) por oclusión incompleta y vocalización que ocurre debido a una pobre coordinación entre las funciones laríngeas y supralaríngeas.

1.3.1 Función Velofaríngea.

Los cambios en las frecuencias de los formantes, reducción en la amplitud de los formantes, incremento en el ancho de banda de los formantes, presencia de formantes nasales y antiformantes caracterizan acústicamente la nasalidad [17].

1.3.2 Coordinación de la actividad laríngea y supralaríngea.

El VOT se define como el intervalo entre la oclusión y el comienzo de energía periódica, el intervalo fisiológico entre la liberación de la constricción de consonante y el comienzo de la vibración de las cuerdas vocales. Es el índice más usado para evaluar la coordinación entre estos subsistemas, aunque debe ser combinado con otras medidas para emitir un criterio, por ejemplo el LGD (laryngeal devoicing gesture) que está compuesto por Tiempo de Oclusión+VOT.

Se ha demostrado que el intervalo de VOT cumple las siguientes condiciones:

VOT (/b d g/) < VOT (/p t k/)

VOT (/b /) < VOT (/d/) < VOT (/g/)

VOT (/p/) < VOT (/t/) < VOT (/k/)

El lenguaje Español tiene VOT negativos para las plosivas sonoras, mientras que en Inglés son mayormente positivos. Las mujeres producen mayores VOT en plosivas sordas y en los niños cambia ligeramente con la edad. Cuando la plosiva es seguida por la vocal /i/ el VOT es mayor que para la vocal /a/ [19]. Un incremento en la razón del habla causa un decremento del VOT para plosivas sordas. Las plosivas sordas producidas con altos valores de F0 producen VOT más cortos que con valores medios o bajos [20] y producen menores tiempos en habla fluida y lectura que en palabras.

1.3.3 Duración de la fase de explosión.

También existe una mayor duración de la barra en las oclusivas sordas.

(7,39 mseg. sordas y 5,83 mseg. sonoras (s ,002; f=9,545)).

En 1998 [1] reportan los siguientes datos para las oclusivas sordas, 5,6 mseg para [p], 7 mseg para [t] y 5 a 6 mseg para [k], este último valor aumenta hasta 14 mseg cuando toman en consideración las varias barras de explosión reportadas por los autores.

Por su parte, Martínez Celdrán [21] apunta que la explosión tiene una gran importancia en lo concerniente al punto de articulación. Se han hecho estudios perceptivos y se ha determinado que en el caso de las oclusivas bilabiales la ausencia de explosión no interfiere para la percepción de la oclusiva, sin embargo en el caso de las otras consonantes la ausencia lleva a la confusión del lugar de articulación.

1.4 Ejercicio de Diadocoquinesis

Diadocoquinesis: Se define como la habilidad de realizar movimientos musculares alternativos.

El ejercicio de Diadocoquinesis es la repetición de segmentos articulatorios que prueban la habilidad de realizar movimientos articulatorios alternativos.

Uno de los usos más comunes de este ejercicio es en la práctica clínica, sus orígenes se establecen en una serie de estudios llevados a cabo en las Clínicas Mayo en los finales de la década de los 60's y principios de los 70, Darley, Aronson y Brown ([27];[29];[30];[31];[32];[33]; [34]) examinaron las características perceptuales del habla de grupos de disártricos. Los resultados de estos estudios y de trabajos relacionados estrechamente con ellos, se consideran aun hoy las bases del diagnóstico diferencial clínico de las disartrias.

Cada paciente ejecutó tres actividades habladas que clínicamente han demostrado (según las clínicas Mayo) que proveen un máximo de información en el menor tiempo posible:

1. Una muestra del lenguaje contextual obtenida poniendo al paciente a leer en voz alta párrafos estándar y permitiendo al evaluador medir la forma en que el paciente integra las variadas características fonatorias, resonatorias y prosódicas del lenguaje o discurso contextual.
2. Prolongación de la vocal /a/ con el propósito de permitir al evaluador de juzgar la calidad, amplitud, duración y persistencia del control fonatorio.
3. Repeticiones de los sonidos /pa/, /ta/, /ka/ (Diadocoquinesis), tan rápido y parejo como sea posible con el objetivo de proveer información acerca del ritmo, regularidad, y duración de todo tipo de movimientos articulatorios.

Basándose en estos tres ejercicios, un panel de tres especialistas evaluó auditivamente la percepción de 38 dimensiones subjetivas, en una escala de 7 que aun se utilizan en el análisis acústico de las disartrias.

1.5 Algoritmos de detección de plosivas.

El *procesamiento general* del algoritmo de Liu [22], consiste en el cálculo de un espectrograma y su división en seis bandas de frecuencia, luego se realizan dos pasos de procesamiento (fino y grueso). 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 (ROR, derivada de la

energía) y los picos de las derivadas son detectados. Estos picos representan la posición en el tiempo de cambios espectrales abruptos en cualquiera de 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 4. se representa el diagrama en bloques correspondiente.

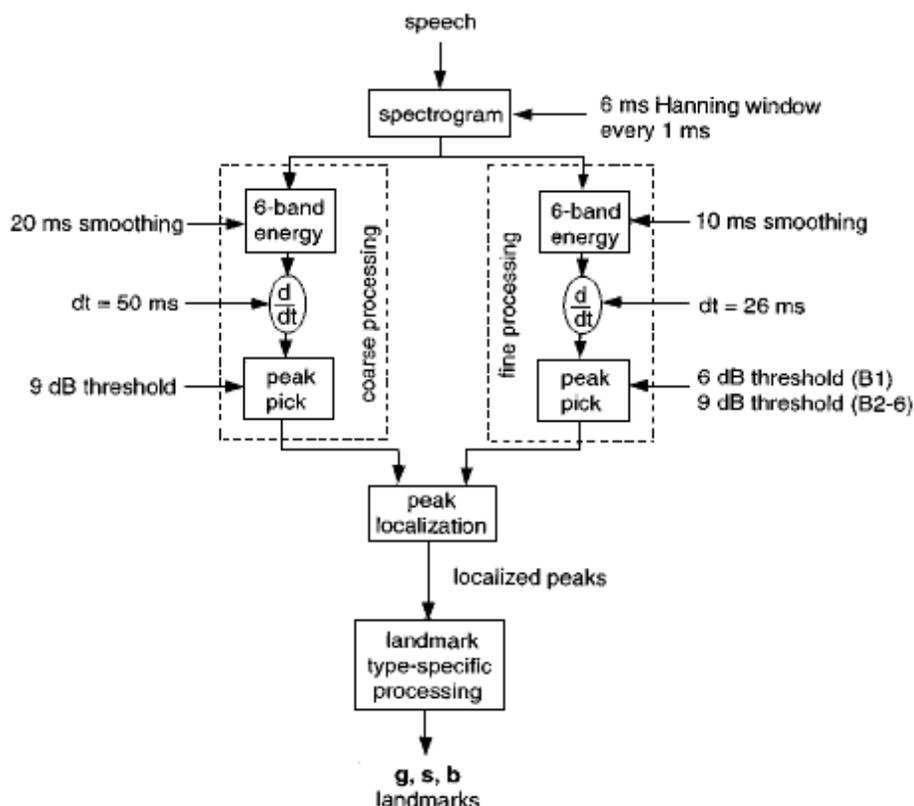


Figura 4 DIAGRAMA DE BLOQUES DEL ALGORITMO DE LIU, 1996.

Un espectrograma de banda ancha es calculado con una ventana de Hanning de 6ms cada 1ms. Cada trama de 6ms es rellenada con ceros hasta alcanzar 512 puntos. El espaciamiento entre los puntos para la transformada discreta de Fourier es de 31.2 Hz para que las amplitudes de los picos espectrales puedan estar bien estimadas. 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.

1.5.1 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 que debe ser como mínimo 20 ms, que es la duración mínima de una vocal promedio. 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 BANDA1 que se rige por un umbral de 20 dB de acuerdo con [37].

1.5.2 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. 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.

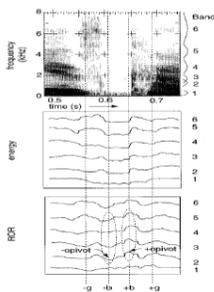


Figura 5 Detección De Marcas Burst Según El Algoritmo De Liu

Los resultados que ofrece este algoritmo se resumen en la Tabla 1. Donde E_1 no incluye los errores de inserción y E_2 no incluye los errores de substitución.

Tabla 1

Landmark type	No. Tokens	Del	Subs	Ins	Neut	E_1	E_2
g(lottis)	1052	1%	3%	7%	0%	11%	8%
s(onorant)	332	10%	2%	21%	9%	33%	31%
b(urst)	213	5%	1%	6%	33%	12%	11%
Total	1597	4%	2%	10%	7%	16%	14%

1.5.3 Detector c (oclusive)

El método propuesto por Salomon [35] es una evolución del método de Liu de dividir la señal en bandas y buscar umbrales de energía. Además incluye información de la frecuencia fundamental, de periodicidad y aperiodicidad de la señal.

Utiliza un banco de filtros gammatone de 60 canales con frecuencias características basadas en la escala ERB [23]. La envolvente de la energía se extrae mediante la transformada de Hilbert según la Ecuación 4.

$$e_i(t) = |x_i(t) + j \cdot H\{x_i(t)\}|$$

Ecuación 4

Las medidas de energía y periodicidad se extraen a partir del cálculo del AMDF [24] y el uso del algoritmo de convex hull [25]. Las sumatorias de energías periódicas y aperiódicas por canales y tramas se utilizan en la estimación de las fronteras entre segmentos con actividad vocal (v), consonantes sonoras (s) y consonantes oclusivas (c). Los resultados generales para las oclusivas sordas son 9.14% no detectadas o substituidas en cierre y oclusión y 1.03% respectivamente en cierre y liberación.

1.6 Consideraciones finales.

Existe una gran cantidad y variedad de estudios realizados en el campo de la fonética acústica, buena parte de los cuales se basan en el análisis de los primeros formantes de la voz. Gracias a estos estudios conocemos parte de las características espectrales más importantes de los sonidos que intervienen en la comunicación oral.

Un adecuado conocimiento de la posición y evolución de los formantes del habla, facilita considerablemente la comprensión de los espectros de voz.

Los estudios que se realizan sobre fonética acústica, ayudan de una manera muy importante al diseño y realización de aplicaciones de tratamiento de la voz, especialmente en los campos de reconocimiento y síntesis del habla.

La resolución de los diversos objetivos que plantea el tratamiento automático de la voz, obliga a la creación de equipos de investigación de naturaleza multidisciplinar.

El campo cubierto por la fonética acústica entra de lleno en las áreas que deben ser cubiertas por estos equipos.

Debido a la variabilidad que existe entre las diferentes realizaciones del habla, resulta imposible establecer unas reglas deterministas de caracterización espectral de sonidos. Sin embargo, conocemos las pautas que definen la posición y evolución de los formantes.

Los fenómenos de coarticulación dificultan el análisis de la evolución de los formantes, aunque también nos muestran parte de los principios básicos sobre los que se asienta la comprensión del lenguaje oral.

El campo de la fonética acústica requiere un enfoque experimental sobre el que se asienten las teorías que los investigadores desarrollan. Por ello, resulta muy importante desarrollar métodos y herramientas que faciliten la creación de aplicaciones que muestren con claridad las distintas características espectrales del habla humana.

Conclusiones del capítulo. Luego de una meticulosa búsqueda, que incluyó medidas relacionadas con las consonantes plosivas y la articulación en pacientes normales, queda demostrada la necesidad de una herramienta que detecte consonantes plosivas. Para la realización de esta herramienta se considera que las plosivas presentan un punto de cierre y arranque de las cuerdas vocales.

CAPÍTULO 2. Materiales y Métodos

Resumen: En el desarrollo de este capítulo se definen los materiales y métodos utilizados para la realización del presente trabajo. Se explica las características y composición de la muestra de pacientes, así como las distintas herramientas utilizadas. Se realiza una descripción del algoritmo diseñado, 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 Diseño del algoritmo

Con el propósito de crear un algoritmo para detectar plosivas en el ejercicio de Diadocoquinesis se realizaron una serie de modificaciones en dos algoritmos referenciados en la literatura [22] [35]. Para llevar a cabo esta tarea se utiliza información referenciada e información empírica recogida a lo largo del trabajo con la base de datos seleccionada.

Se toma el procesamiento general del algoritmo de Liu en este caso con un aumento en el número de bandas; para ello se filtra la señal con un banco de filtros Gammatone [38] con frecuencias centrales (CF) desde 100Hz hasta la mitad de la frecuencia de muestreo, este banco de filtros constituye uno de estos modelos de procesamiento de voz que se basan en simular el funcionamiento de la cóclea y esto provee al algoritmo desarrollado de un acercamiento al funcionamiento de la audición humana, ósea al juicio de los expertos.

Existen factores que hacen del banco de filtro Gammatone un sistema con buen desempeño, estas razones son: *Fisiológicas*, *Psicológicas* y por último *Práctica*.

1. *Fisiológicas:* La respuesta de impulso de la función de filtro de Gammatone suministra una compatibilidad excelente con la respuesta de impulso obtenida mediante la técnica

de *revcor* (Correlación entre los datos de una fibra auditiva principal y la forma de onda recibida por el oído) [39].

El filtro Gammatone es definido en el dominio de tiempo por la respuesta de impulso que se describe en Ecuación 5:

$$Gt(t) = at^{(n-1)} \exp(-2pbt) \cos(2pf_0t +) \quad (t > 0)$$

Ecuación 5

Dónde:

n: orden del filtro y este valor determina la sintonía o *Q*, del filtro en gran parte.

*f*₀: frecuencia central

b: Determina la longitud de duración de la respuesta al impulso y por lo tanto del ancho de banda del filtro Gammatone

a: constante real de normalización

p: constante π

2. *Psicológicas*: Las propiedades de selectividad en frecuencia, mediadas fisiológicamente en la cóclea y medidas psicofísicamente en seres humanos, son convergentes. Primero, la característica de magnitud del filtro GT de orden 4 es muy similar a la función *roex*, comúnmente usada para representar la característica de magnitud del filtro auditivo humano [40]; [41]; [42]. En segundo lugar, el estudio realizado por [43] indica que los anchos de banda del filtro auditivo tienen una distancia fija en la membrana basilar. En este estudio, específicamente, se han resumido los datos humanos sobre el ancho de banda rectangular equivalente (ERB) al filtro auditivo humano como:

$$ERB = 24.7(4.37f_c/100 + 1)$$

Ecuación 6

Las ecuaciones Ecuación 5 y Ecuación 6 juntas definen al banco de filtro Gammatone si se incluye la suposición común de que las frecuencias centrales del filtro están distribuidas a través de la membrana basilar en proporción a su ancho de banda. Cuando f_c/b es grande como en el caso auditivo, el ancho de banda del filtro es proporcional a b y la constante de proporcionalidad solo depende del orden del filtro, n . Cuando el orden es 4, b es $1.019ERB$. A los $-3bB$ el ancho de banda del filtro Gammatone es 0.887 veces el ERB . Para concluir, el GT es un filtro de fase mínima y, aunque la fase característica del filtro auditivo no se conoce, resulta razonable suponer que está cerca a la fase mínima.

3. *Práctica:* Un análisis en el dominio de la frecuencia del filtro GT revela que un n -ésimo orden del filtro GT puede ser aproximado por una cascada de n filtro idénticos de primer orden y el filtro de primer orden GT digital tiene una recursividad muy eficiente. El banco de filtro auditivo Gammatone resultante suministra una buena compensación entre la exactitud con la que simula el filtrado coclear y la carga computacional.

El inventariado de la señal filtrada, la transformada de Fourier y los procesamientos de energía se realizan de manera similar a la propuesta de Liu. Los máximos y mínimos en el ROR se seleccionan a partir de $\pm 9dB$ respectivamente y espaciados a $40ms$ que es el tiempo mínimo entre señales acústicamente abruptas [22] Figura 6.

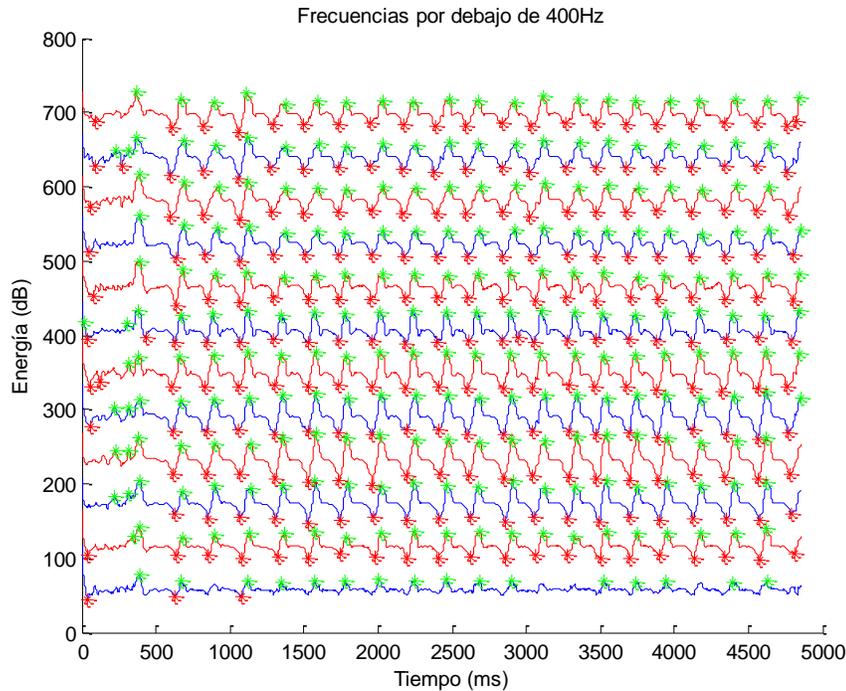


Figura 6 ROR de la señal kdtb.wav procesada

2.1.1 Detector de actividad Glotal

Debido a que el ejercicio de Diadoquesis produce numerosos segmentos de actividad glotal que están precedidos por la consonante oclusiva en estudio; es necesario implementar un detector que localice el segmento sonoro de la sílaba y permita la posterior detección del inicio consonante plosivas. Para ello se realizan modificaciones a la propuesta del Detector **g**.

1. Localizar todos los canales cuyas frecuencias centrales estén por debajo de 400Hz y a continuación con la ayuda de un detector grueso de frecuencia fundamental (F0) [36] calcular el valor medio de F0.
2. Proceder a eliminar del proceso de detección a todas las bandas con frecuencias centrales inferiores a F0. De esta manera se maximiza la posibilidad de detectar un cambio abrupto, ya que se utiliza solamente las bandas que poseen armónicos de la señal de voz y que no contienen cambios espectrales significativos relacionados con las consonantes Figura 7.

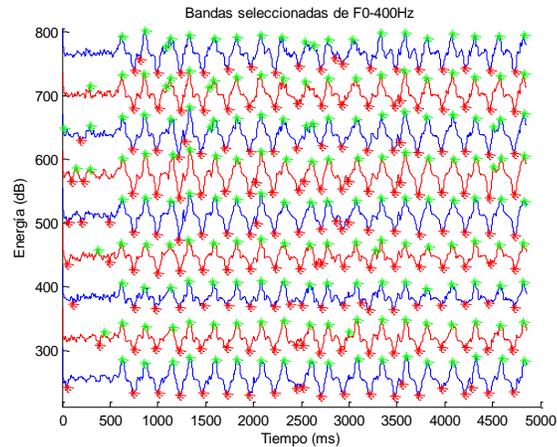


Figura 7 Bandas de ROR seleccionadas para la detección.

3. Para evitar los errores que introducen los picos procedentes de ruidos se calcula un umbral adaptativo de energía que se actualiza en cada detección, el mismo se inicializa a 20dB por debajo del máximo de la Banda 1 del procesamiento grueso de Liu calculado para la señal en estudio ya que la energía de los canales seleccionados en el paso anterior son redundantes y presentan menos variabilidad Figura y Figura 9. La actualización del umbral consiste en determinar el valor medio de energía en dicha banda para los alrededores de cada marca g detectada, estos se encuentran delimitados por la marca g positiva y el primer mínimo en el ROR en un intervalo mayor de 20ms.

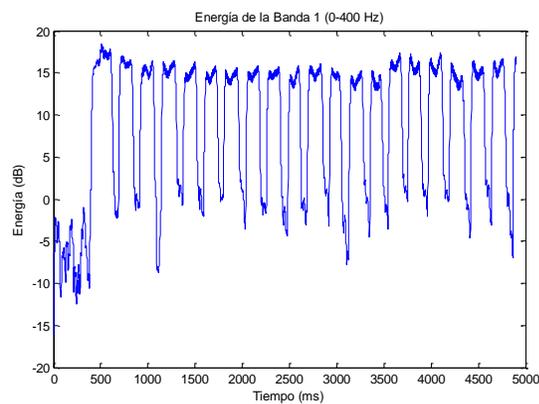


Figura 8 Energía de la Banda 1 0-400Hz de la señal kdtb.wav procesada

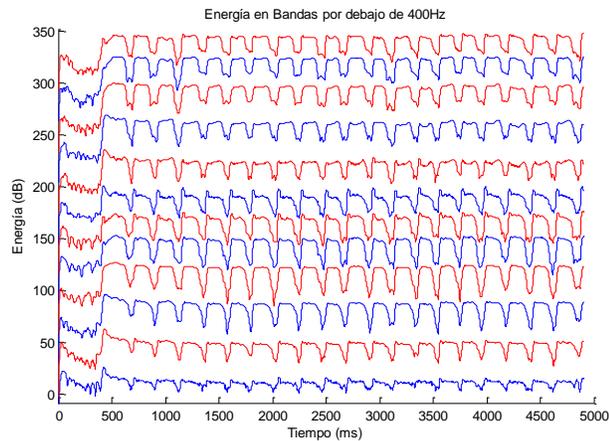


Figura 9 Energía de las Bandas de F0 a 400 Hz de la señal kdtb.wav procesada

4. Durante la detección todos los candidatos a marca g con energías inferiores al umbral son eliminados.
5. Por último se busca que exista una coincidencia en tiempo de marcas g positivas en al menos la mitad de las bandas seleccionadas, de no cumplirse se elimina el candidato a marca g positiva.
6. Con el propósito de mejorar la precisión temporal se localiza la marca g en el procesamiento fino utilizando una ventana de 30ms alrededor de la marca g del procesamiento grueso.
7. Para la detección de la marca g negativa se cuenta con el arreglo de marcas g positivas y la energía de la Banda 1. Se extraen los picos negativos a partir de 20ms después una marca g positiva y 20ms antes de la siguiente. A este conjunto se les hace un ordenamiento de acuerdo con los valores de energía asociados a los mismos en la Banda 1 y se toma el menor del conjunto que se asocia con el fin de vibración de las cuerdas vocales Figura 10.

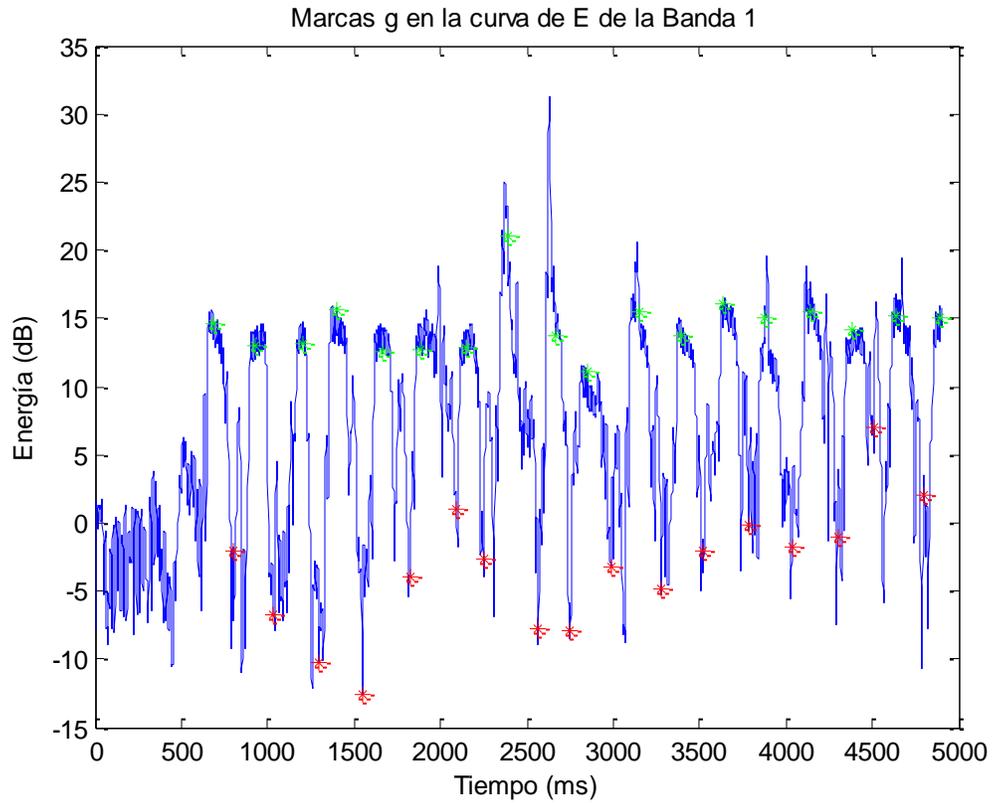


Figura 10 Curva de energía de la Banda 1 grabación kcampillo.wav

8. Con el propósito de mejorar la precisión temporal se localiza la marca g en el procesamiento fino utilizando una ventana de 30ms alrededor de la marca g del procesamiento grueso.

Para corroborar el desempeño del detector se colocan las marcas g en el espectro de banda estrecha de la señal Figura 11.

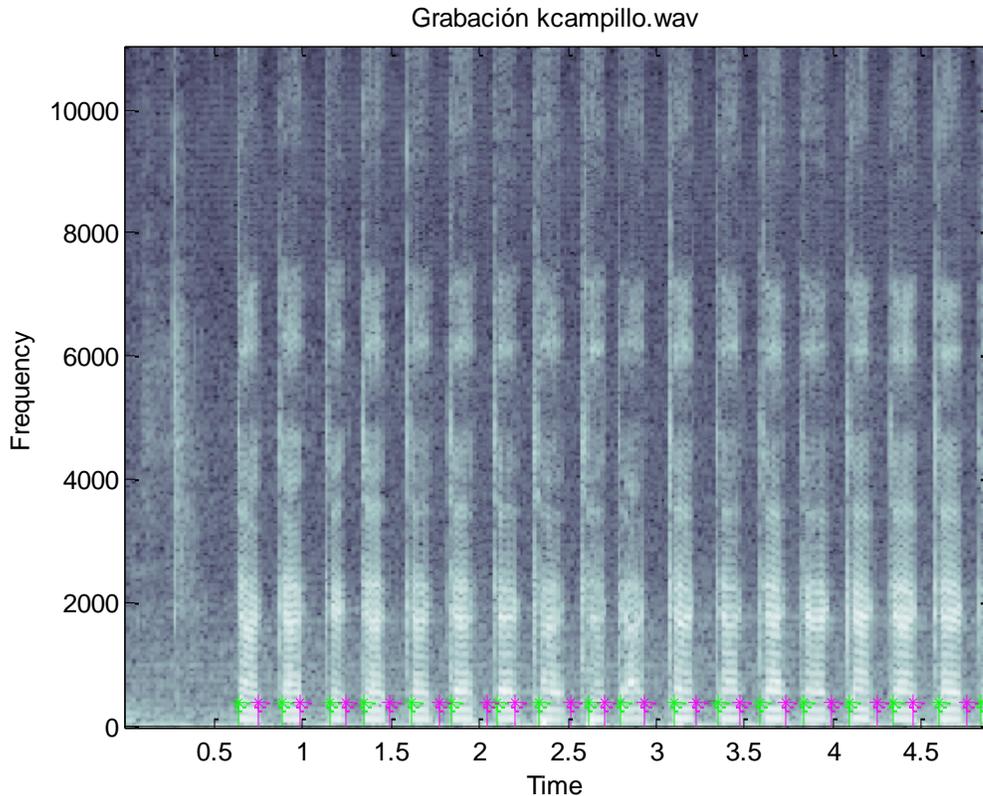


Figura 11 Espectrograma de banda estrecha de la grabación kcampillo.wav. Asteriscos marcas g

2.1.2 Detector de plosivas (b)

Para la detección de plosivas se localizan todas las bandas con CF comprendidas entre 800 y 5000Hz Figura . Luego se realiza un barrido en una secuencia de segmentos de tiempo limitados por una marca g negativa y la marca g positiva siguiente o en caso de no existir marca g negativa se limita a 100ms antes de la marca g positiva. Todos los picos detectados en las bandas del ROR deben exceder un umbral empírico de energía fijado a 5dB por encima del valor medio de la energía de la banda correspondiente al pico en cuestión.

La ubicación del inicio de la oclusiva se asocia a la posición del pico con mayor valor de energía detectado en un segmento. Estos se grafican en el espectrograma de banda estrecha junto con las marcas g Figura 13.

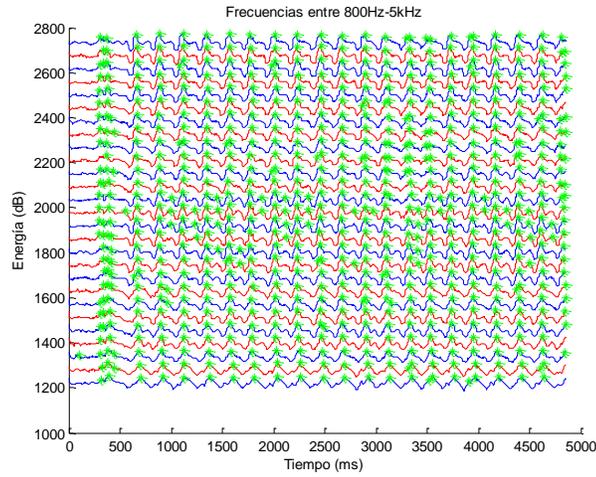


Figura 12 Bandas de ROR seleccionadas para la detección de plosivas.

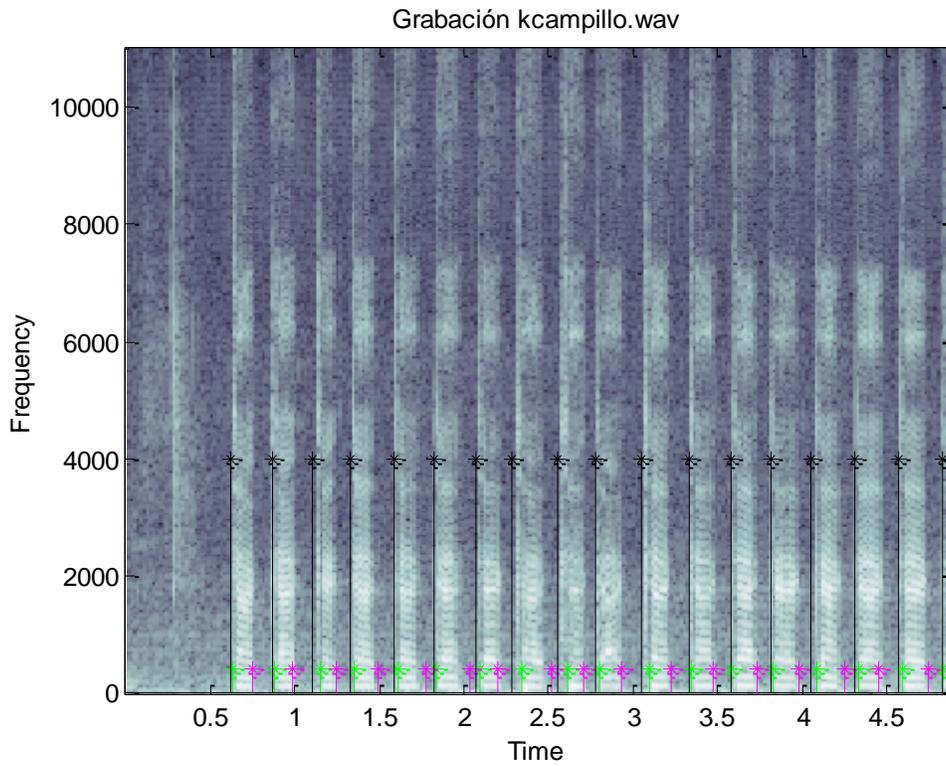


Figura 13 Espectrograma de banda estrecha de la grabación kcampillo.wav asteriscos negros posición de las plosivas, otros asteriscos marcas g

2.2 El experimento

En la base de datos utilizada se cuenta con 18 sujetos sanos, 10 mujeres y 8 hombres. Cada sujeto realiza 3 grabaciones con los ejercicios de Diadocoquinesis /pa/, /ta/ y /ka/ durante 5 segundos pronunciando el mayor número de repeticiones posibles del ejercicio. Las grabaciones se capturan con una frecuencia de muestreo $f_s=22050\text{Hz}$ y una resolución de 16bits.

Se procesa cada grabación de forma independiente y se evalúa el desempeño del detector contando las plosivas detectadas correctamente, el total de plosivas existentes (evaluación subjetiva) y las inserciones. Se calcula el error absoluto en porcentaje mediante Ecuación 7.

$$e = \frac{|plosivas - detectadas|}{plosivas} 100\%$$

Ecuación 7

2.3 Conclusiones del capítulo.

El algoritmo propuesto ha logrado detectar las consonantes plosivas en el ejercicio de Diadocoquinesis de la muestra seleccionada, un ejemplo del desempeño del mismo sobre un paciente se encuentra en el Anexo I, con este experimento se ha recolectado una valiosa información que se discute en el próximo capítulo.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En el presente capítulo se discuten los resultados obtenidos para cada sujeto en cada una de las consonantes /pa/, /ta/ y /ka/ sin tener en cuenta el género. Se tabulan los resultados y se emiten conclusiones acerca del desempeño del detector.

3.1 Consonante /ka/.

En la Tabla 2 se muestran los resultados obtenidos por cada paciente en la pronunciación de la sílaba /ka/, donde se obtiene un porcentaje de error significativamente bajo que manifiesta una buena detección de las plosivas del algoritmo implementado.

Tabla 2 Pronunciación de la consonante /ka/.

Pacientes	Plosivas	Plosivas detectadas	Plosivas no detectadas	Error total (%)
Betyg	25	25	0	
Betyp	22	22	0	
Campillo	18	18	0	
Drianet	20	20	0	
Ds	25	23	2	
Dtb	31	31	0	
Elaine	23	21	2	
Iroel	21	19	2	
Kappa	34	34	0	
Mariae	22	22	0	
Mariav	15	14	1	
Meily	24	23	1	
Orozco	20	20	0	
Rene	34	32	2	
Robert	20	20	0	
Una	15	15	0	
Vento	19	19	0	
Yisel	22	22	0	
Total	410	400	10	2.5

En los pacientes Ds, Elaine, Iroel y Rene, el algoritmo deja de detectar dos plosivas y es debido a fallos en el detector g mayormente por malas oclusiones durante el ejercicio, esto no es significativo para el funcionamiento del algoritmo.

Por ejemplo en la Figura 14, en el intervalo de tiempo 2.5 y 3 segundos el algoritmo detecta como un único segmento de actividad glotal porque no existe un cierre correcto de las cuerdas vocales.

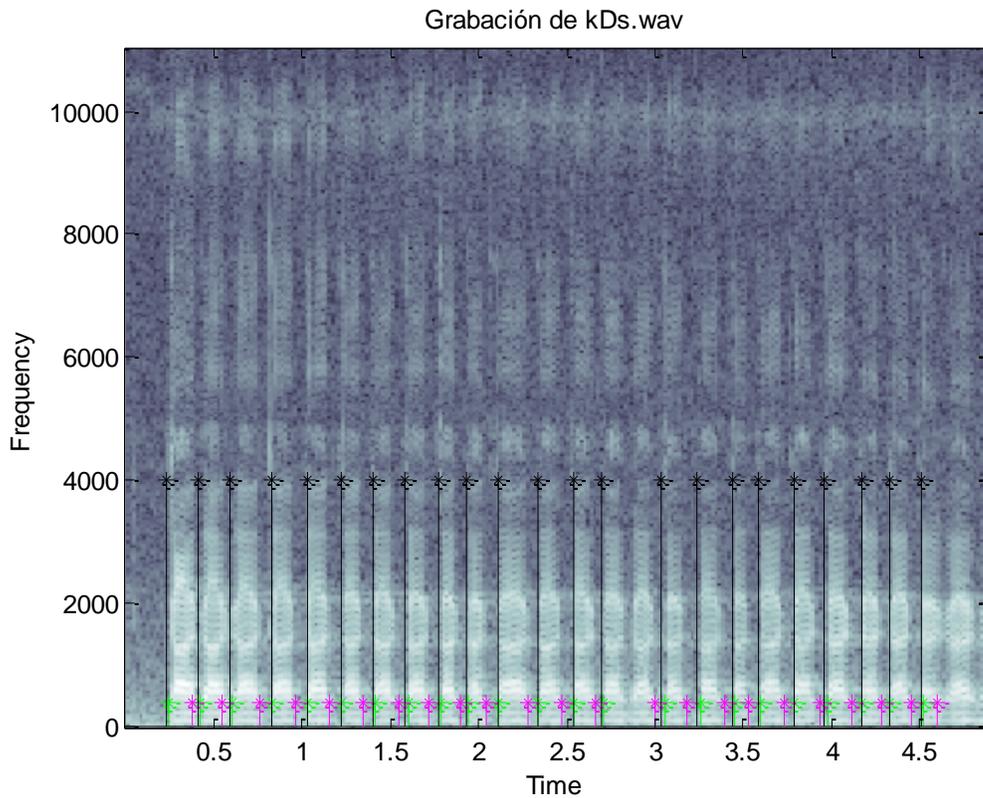


Figura 14 Espectrograma de banda estrecha de la grabación kDs.wav asteriscos negros posición de las plosivas, otros asteriscos marcas g

3.2 Consonante /pa/

En la Tabla 3 se muestran los resultados obtenidos por cada paciente. El porcentaje de error del algoritmo se incrementa con la señal Elaine, con 9 consonantes omitidas presuntamente por la baja calidad articulatoria del sujeto de experimentación.

Tabla 3 Pronunciación de la consonante /pa/.

Pacientes	Plosivas	Plosivas detectadas	Plosivas no detectadas	Error total (%)
Betyg	29	29	0	
Betyp	25	24	1	
Campillo	23	21	2	
Drianet	18	18	0	
Ds	35	29	6	
Dtb	24	22	2	
Elaine	29	20	9	
Iroel	24	22	2	
Kappa	32	32	0	
Mariae	26	26	0	
Mariav	25	22	3	
Meily	25	25	0	
Orozco	11	11	0	
Rene	30	27	3	
Robert	28	28	0	
Una	17	16	1	
Vento	18	18	0	
Yisel	30	28	2	
Total	449	418	31	6.9

3.3 Consonante /ta/

En la Tabla 4 se observan los resultados obtenidos por cada paciente en las evaluaciones subjetivas y el detector. El porcentaje de error del algoritmo alcanza un valor ligeramente menor que en la consonante /pa/. Es notable que la mayor cantidad de omisiones se encuentra en la grabación tdrianet.wav, presuntamente por las maloclusiones y las sustituciones de la sílaba /ta/ por /ga/. Además cabe mencionar que la grabación telaine.wav vuelve a mostrar un elevado número de errores.

Tabla 4 Pronunciación de la consonante /ta/.

Pacientes	Plosivas	Plosivas detectadas	Plosivas no detectadas	Error total (%)
Betyg	31	31	0	
Betyp	24	24	0	
Campillo	22	22	0	
Drianet	18	18	0	
Ds	33	18	15	
Dtb	31	31	0	
Elaine	33	28	5	
Iroel	27	27	0	
Kappa	31	31	0	

Mariae	25	25	0	
Mariav	23	22	1	
Meily	25	24	1	
Orozco	32	32	0	
Rene	34	32	2	
Robert	28	27	1	
Una	25	25	0	
Vento	17	17	0	
Yisel	38	37	1	
Total	497	471	26	5.2

3.4 Conclusiones del capitulo

En este capítulo se obtienen resultados satisfactorios con respecto al algoritmo implementado. El error de manera general es de un 4.9% para la base de datos analizada.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

1. El algoritmo diseñado a partir de la idea propuesta por Liu y Salomon permite detectar las consonantes plosivas.
2. Además de cumplir el objetivo propuesto se logran las siguientes medidas: Inicio y fin de vibración de las cuerdas vocales, los espectros de banda ancha y banda estrecha con ventana de hanning de la señal, y la señal descompuesta en 60 canales por un banco de filtros Gammatone. Todas aplicables a posteriores etapas de análisis y detección en la señal de habla.
3. Se introduce el cálculo del error para cada uno ejercicio Diadocoquinesis.

Recomendaciones

1. Con el algoritmo implementado en esta investigación y los resultados obtenidos se han creado las condiciones para una fase de post-procesamiento que efectúe un cálculo más preciso de medidas objetivas para detectar consonantes plosivas en voces patológicas. Para esto se podría utilizar la base de datos de Clínicas de Mayo y establecer las diferencias entre pacientes femeninos y masculinos.
2. Con las medidas calculadas y las marcas encontradas y posicionadas en el espectro se han creado las condiciones para una fase de post-procesamiento que efectúe un cálculo más preciso de medidas objetivas en las dimensiones antes mencionadas (calidad vocal, articulación, nasalidad, prosodia) para alcanzar una medida objetiva de inteligibilidad.

REFERENCIAS BIBLIOGRÁFICAS

- [1]. Borzone, A. M. y J. Gurlekian. 1980. Rasgos acústicos de las consonantes oclusivas
- [2]. Asensi, L., S. Portolés y A. del Río. 1998. Barra de explosión, VOT y frecuencia de las oclusivas sordas del castellano. *Estudios de Fonética Experimental*. 9: 201-220. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética españolas. *Fonoaudiológica*. 26: 326-330. Buenos Aires-Argentina.
- [3]. Castañeda V., M. L. 1986. El VOT de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*. 2: 91-108. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética.
- [4]. Martínez Celdrán, E. 1991a. Sobre la naturaleza fonética de los alófonos de /b d g/ en español y sus distintas denominaciones. *En Verba*. 18: 235-253. España: Universidad de Barcelona.
- [5]. Martínez Celdrán, E. 1991b. Duración y tensión de las oclusivas no iniciales del español: Un estudio perceptivo. *Revista Argentina de Lingüística*. 1: 52-71. Mendoza-Argentina: CONICET.
- [6]. Martínez Celdrán, E. 1992. Un mismo parámetro fonético en el fondo de la lenición protorromance: La duración. *Estudios Filológicos en Homenaje a Eugenio de Bustos Tovar*. 117-136. Salamanca-España: Ediciones Universidad.
- [7]. Martínez Celdrán, E. 1993. La duración de la nasal precedente como índice de la tensión de las oclusivas españolas. *Estudios de Fonética Experimental*. IV: 331-339. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética.

- [8]. Machuca A., M. J. 1997. Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en habla espontánea. Tesis Doctoral. Barcelona-España: Departamento de Filología española. Facultad de Filosofía y Letras. Universidad Autónoma de Barcelona.
- [9]. Mujica Agudo, E., M. Santos Soto y J. Herraiz Moya. Duración de las transiciones en las oclusivas sordas del Castellano. *Estudios de Fonética Experimental*. 4: 103-122. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética.
- [10]. Poch, M. D. 1984. Datos acústicos para la caracterización de las oclusivas sordas del español. *Folia Phonetica*. 1: 89-106. Lleida, España: Revista del Laboratori de Fonètica “Pere Barnils”, Estudi General de Lleida, Universitat de Barcelona.
- [11]. Soler, A y J. Romero. 1999. The role of duration in stop lenition in spanish. *Actas del XIVth International Congress of Phonetic Sciences*. 483-486. San Francisco.
- [12]. Kent, R.D. , Weismer, G. , Kent, J. F. , Vorperian, H.K. & Duffy, J.R. “Acoustic Studies of Disarthric Speech: Method, Progress, and Potential”. *Journal of Communication Disorders*, vol. 32, pp 141-186. 1999.
- [13]. Tjaden, K., & Turner, G.S., “Spectral properties of fricatives in amyotrophic lateral sclerosis”, *Journal of Speech, Language, and Hearing Research*, 36, pp 1177– 1190, 1997.
- [14]. Hernandez – Diaz Huici, M. E., Verhelst, W., “The development of objective tools to assess and manage articulation and prosody in hearing impaired and dysarthric patients”, Research in Brussels RIB-2006/006 Report, August 2007.
- [15]. Hernández – Díaz Huici, M. & Verhelst, W., “Spectral transition features in dysarthric speech”, Proc. of MAVEBA, Firenze, Italy, pp 139- 142, 2007.
- [16]. Ackerman, H., & Ziegler, W., “Articulatory deficits in Parkinsonian dysarthria: An acoustic analysis”, *Journal Neurology, Neurosurgery, and Psychiatry*, 54, pp 1093– 1098, 1991b.
- [17]. Kent, R.D., & Read, W.C., “The Acoustic Analysis of Speech”, San Diego: Singular Publishing Group., 1992.

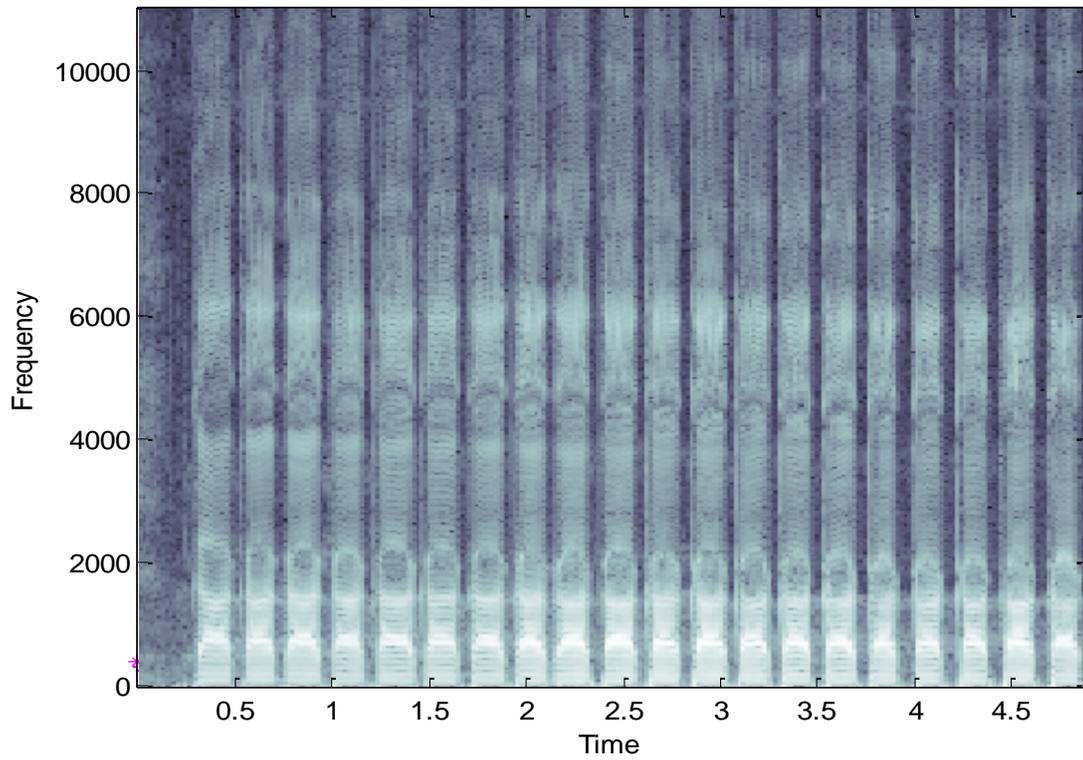
- [18]. Stouten, V., Van hamme, H. “Automatic Voice Onset Time Estimation from Reassignment Spectra”, *Het SPACE-project*. 2009.
- [19]. Whiteside, S., Henry, L., Dobbin, R. Sex difference in voice onset time: A developmental study of phonetic context effects in british english. *J. of ASA*, vol. 116(2), pp 2851-2860. 2004.
- [20]. McCrea, C., Morris, R. The effect of fundamental frequency level on voice onset time in normal adult male speakers. *J. of Speech, Lang and Hearing Res* 48, pp 1013-1024. 2005.
- [21]. Martínez Celdrán, E. 1995. La percepción categorial de /b-p/ en español basada en las diferencias de duración. *Estudios de Fonética Experimental*. 5: 225-239. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética.
- [22]. Liu, S. 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. 1996.
- [23]. B. R. Glasberg and B. C. J. Moore, “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.*, vol. 47, pp. 103–138, 1990.
- [24]. L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [25]. P. Mermelstein, “Automatic segmentation of speech into syllabic units,” *J. Acoust. Soc. Amer.*, vol. 58, pp. 880–883, 1975.
- [26]. Martínez Celdrán, E. 1995. La percepción categorial de /b-p/ en español basada en las diferencias de duración. *Estudios de Fonética Experimental*. 5: 225-239. Barcelona-España: Facultad de Filología, Universidad de Barcelona, Laboratorio de Fonética.
- [27]. Darley, F.L.; Aronson, A.E. & Brown, J.R. “Clusters of deviant speech dimensions in the dysarthria”. *Journal of Speech & Hearing Research*. 12, pp 462-496, 1969.
- [28]. Darley, F.L.; Aronson, A.E. & Brown, J.R. “Differential diagnostic patterns of dysarthria”. *Journal of Speech & Hearing Research*. 12, pp 246-269, 1969.

- [29]. Berry, W.R.; Aronson, A.E.; Darley, F.L. & Goldstein, N.P. "Effects of penicillamine therapy and low-copper diet on dysarthria in Wilson's disease (hepatolenticular degenerative)". *Mayo Clinic Proceedings*. 49, pp 405-408. 1974.
- [30]. Berry, W.R.; Darley, F.L.; Aronson, A.E. & Goldstein, N.P. "Dysarthria in Wilson's.
- [31]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Differential diagnostic patterns of dysarthria". *Journal of Speech & Hearing Research*. 12, pp 246-269, 1969.
- [32]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Disorders". Philadelphia. Saunders. 1975.
- [33]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Signs in Neurologic Disease". *Medical Clinics of North America*, 52, 835-844. 1968.
- [34]. Darley, F.L.; Brown, J.R. & Goldstein, N. "Dysarthria in multiple sclerosis". *Journal of Speech & Hearing Research*. 15, pp 229-245, 1972.
- [35]. O. Deshmukh and C. Espy-Wilson, "Detection of periodicity and aperiodicity in speech signal based on temporal information," in *Proc. 15th Int. Congr. Phonetic Sciences*, Barcelona, Spain, 2003, pp. 1365–1368.
- [36]. Ross, M. J., Shaffer, H.L., Cohen, A., Freudberg, R. & Manley, H. J (1974) "Average magnitude difference function pitch extractor" *IEEE Transactions on Acoustics, Speech and Signal Processing*. 5. 353-362.
- [37]. Stevens, K. N. Prosodic influences on glottal waveform: preliminary data, In *Proceedings of the International Symposium on Prosody*. Japan Society for the Promotion of Science, Yokohama, Japan, pp. 53–64. 1994.
- [38]. Patterson, R.D. & Holdsworth, J., 1993. A FUNCTIONAL MODEL OF NEURAL ACTIVITY PATTERNS AND AUDITORY IMAGES. *Advances in Speech, Hearing and Language Processing*, 3(Parte B), pp.547-63.
- [39]. Boer, E.d. & Jongh, H.R.d., 1978. On cochlear encoding: potentialities and limitations of the reverse-correlation technique. *J. Acoust. Soc. Am*, 63, pp.115-35.
- [40]. Schofield, D., 1985. Visualisations of speech based on a model of the peripheral auditory system. *NPL Report DITC 62/85*.

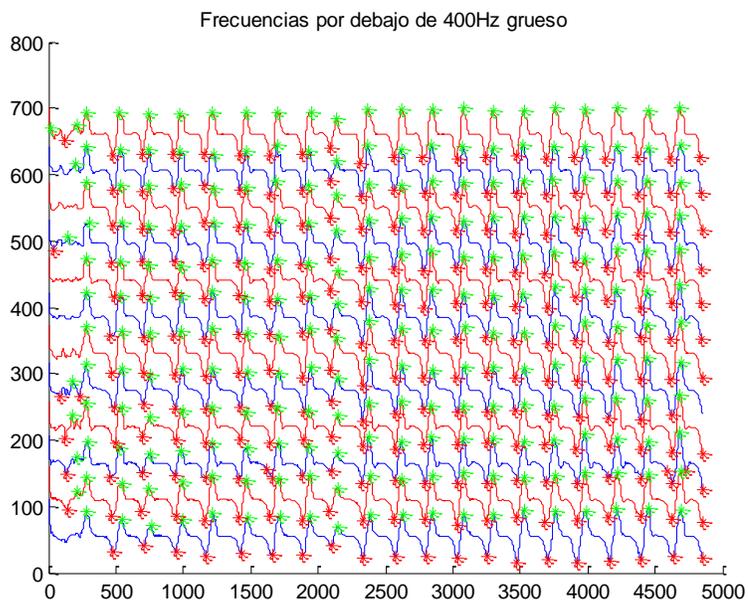
-
- [41]. Patterson, R.D. & Moore, B.C.J., 1986. Auditory filters and excitation patterns as representations of frequency resolution. *Frequency Selectivity in Hearing*, pp.123-77.
- [42]. Glasberg, B.R. & Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47, pp.103-38.
- [43]. Patterson, R.D., Holdsworth, J., Nimmo-Smith, I. & Rice, P., 1988. SVOS Final Report: The Auditory Filterbank. *APU report*, 2341.

Anexo I Ejemplo de corrida para un paciente.

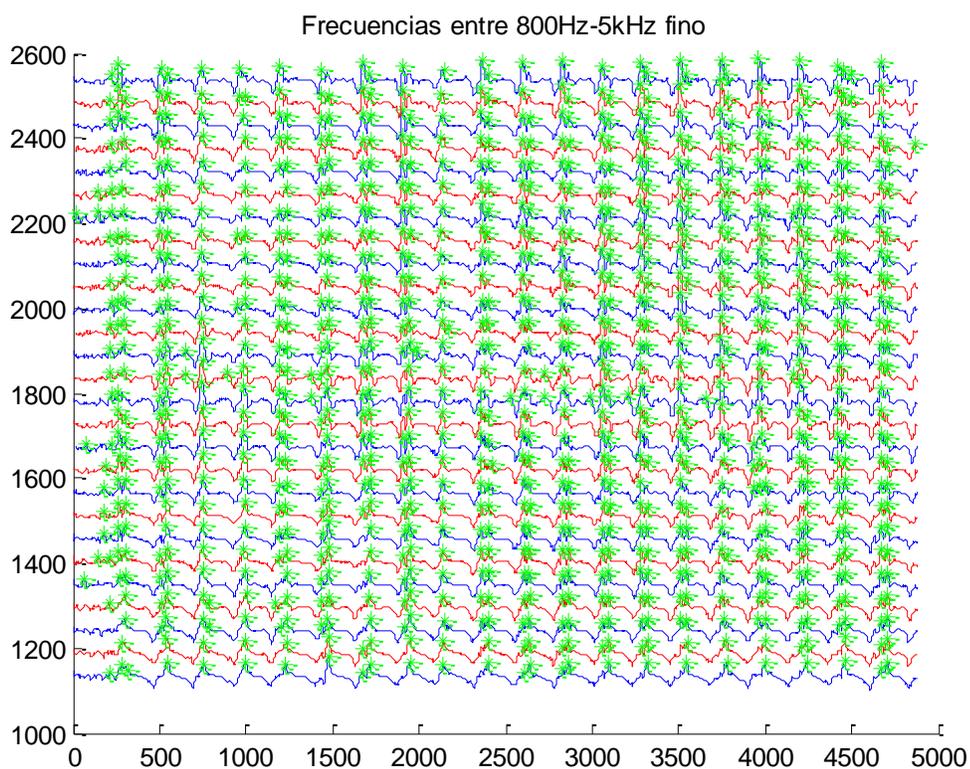
Señal *Orozco.wav*, paciente masculino de 40 años.



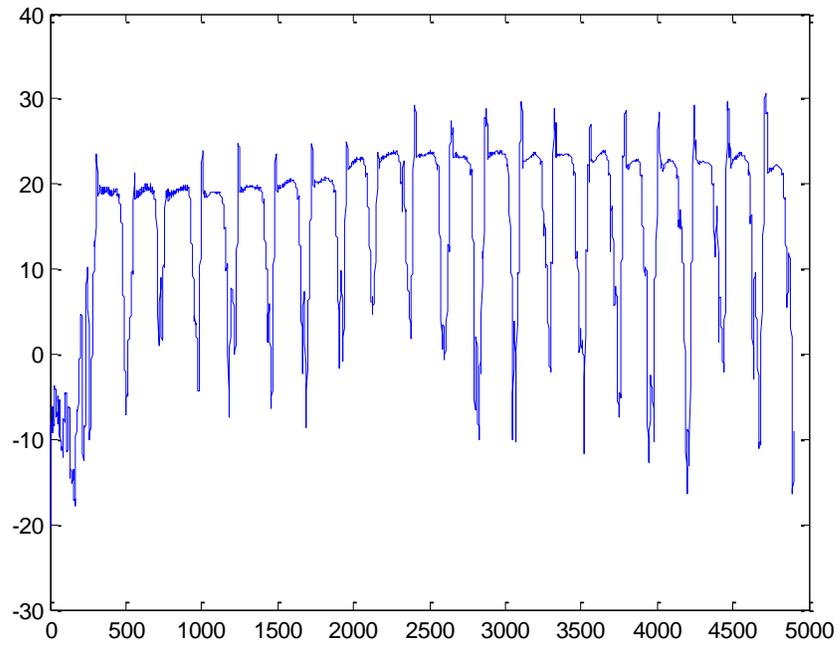
Se puede apreciar en el espectrograma, que están bien definidas todas las bandas de la señal del habla.



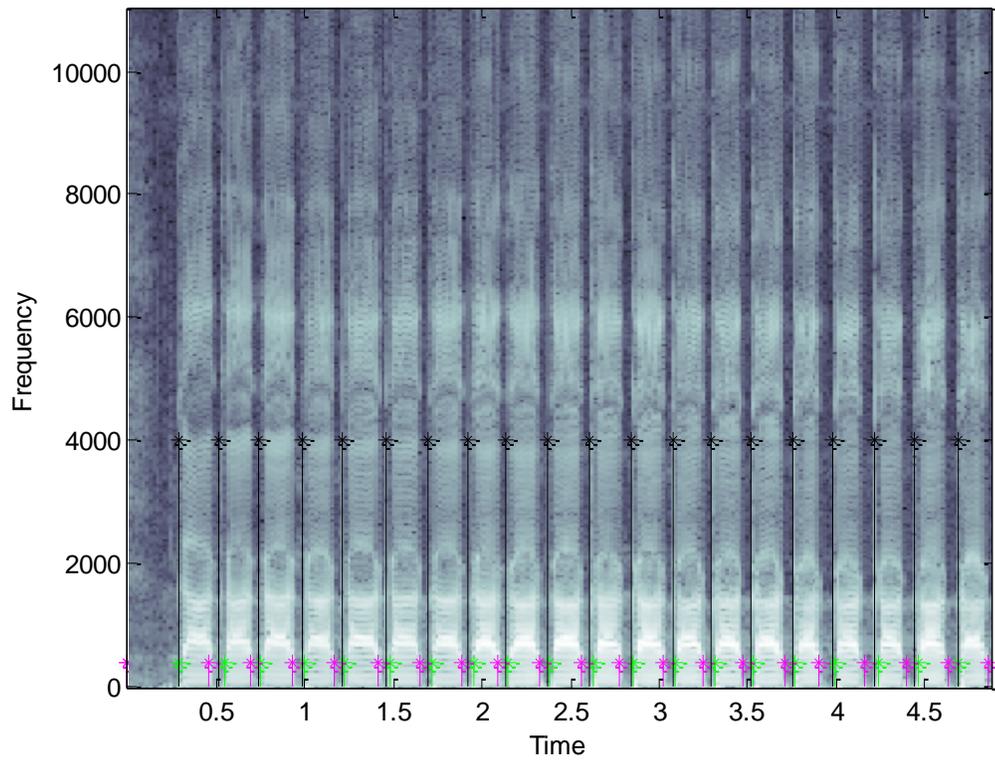
Frecuencias por debajo de 400Hz. Procesamiento grueso.



Frecuencias entre 800Hz y 5kHz. Procesamiento fino.



Energía Total de la señal.



Se muestra como el algoritmo detecta todas las plosivas durante las repeticiones de la sílaba /ka/.