



UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS
VERITATE SOLA NOBIS IMPONETUR VIRILIS TOGA, 1948

FACULTAD DE INGENIERÍA ELÉCTRICA

Centro de Estudios de Electrónica y Tecnología de la Información



TRABAJO DE DIPLOMA

“El modelo auditivo incorporado a la detección de entonación en voces patológicas”

Autor: Drianet Contino Rodríguez

Tutor: Dra. María E. Hernández – Díaz Huici

Santa Clara

2009

"Año del 50 Aniversario del Triunfo de la Revolución"

CON SU ENTRAÑABLE TRANSPARENCIA



Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnología de la Información



TRABAJO DE DIPLOMA

“El modelo auditivo incorporado a la detección de entonación en voces patológicas”

Autor: Drianet Contino Rodríguez

Tutor: Dra. María 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

2009

"Año del 50 Aniversario del Triunfo 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 Tutor

Firma del Jefe de
Departamento donde se
defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

PENSAMIENTO

“Mientras una parte de lo que percibimos penetra a través de nuestros sentidos a partir del objeto que tenemos ante nosotros, otra parte (y tal vez sea la mayor) surge siempre de nuestra propia mente.”

William James

DEDICATORIA

A mis abuelos Victoria y Ernesto, cuyos sueños hoy se ven realizados, por tantos años de esfuerzos en pos de mi formación como persona, por aconsejarme y alentarme en todo momento.

AGRADECIMIENTOS

A mi madre, a mi hermano y a mi padrastro por ser guías inspiradoras.

A mi tutora por su paciencia y dedicación. Sus enseñanzas y consejos han sido vitales tanto para la confección de este trabajo como para mi vida futura.

A Mary por todo su cariño y por brindarme el apoyo necesario en cada momento.

A Alexander por toda ayuda prestada en la confección del trabajo.

A todos mis maestros y profesores que de una forma u otra han aportado su granito de arena en mi formación como profesional.

A todos mis amigos.

A todos aquellos que han contribuido a que este trabajo sea una realidad.

TAREA TÉCNICA

- Revisión bibliográfica para familiarizarse con los conceptos relacionados con la prosodia y dentro de ella conocer acerca de las mediciones de entonación de la voz; además de adquirir conocimientos acerca de los modelos de producción del habla y de los algoritmos para la detección de periodicidad utilizados en la literatura.
- Dominar el uso y las facilidades que brinda el MATLAB ver 7.0 en el procesamiento digital de señales acústicas.
- Programar los siguientes algoritmos:
 - Subharmonic Summation (SHS).
 - SIFT.
 - CEPSTRUM.
 - Autocorrelación.
- Validar los algoritmos programados aplicados a voces patológicas y comparar los resultados obtenidos.
- Realizar el informe de la investigación.

Firma del Autor

Firma del Tutor

RESUMEN

En el estudio de la prosodia es de vital importancia el análisis de la entonación de los hablantes. Se han desarrollado un gran número de algoritmos para la detección de periodicidad de la voz, cuya precisión está dada de acuerdo a los propósitos para su aplicación.

El modelo auditivo en las mediciones de entonación tiene gran desarrollo en el ámbito de la música con una gran muestra de sus resultados publicados en la literatura. Su robustez se puede tener en cuenta a la hora de la modelación prosódica.

En el presente trabajo se realiza un estudio de diferentes detectores de la frecuencia fundamental con el objetivo de encontrar el más robusto y preciso a la hora de obtener la entonación del habla en pacientes que presentan disartrias. La necesidad de la búsqueda de mediciones objetivas para correlacionarlas con las mediciones subjetivas es una motivación para realizar esta tarea, puesto que constituye una herramienta valiosa que puede ser utilizada en el diagnóstico médico y en el tratamiento del paciente.

Los resultados presentan un análisis comparativo de los algoritmos donde se puede determinar el que realiza una detección más eficiente de los contornos de entonación en voces patológicas, en el contexto de una oración, donde existe una mayor complejidad en las mediciones y exige de cierto rigor para el análisis de los resultados.

TABLA DE CONTENIDOS

PENSAMIENTO	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
TAREA TÉCNICA	iv
RESUMEN	v
INTRODUCCIÓN	1
CAPÍTULO 1. MEDIDAS DE ENTONACIÓN	4
1.1. Problemas de las mediciones en voces patológicas	4
1.2. Afecciones en el habla.	5
1.3. Análisis acústico en la caracterización de disartrias.	6
1.1. Prosodia	7
1.4.1. Entonación.	7
1.4.2. Problemas en los contornos de entonación.	8
1.5. El modelo auditivo.	9
1.6. Selección del algoritmo.	10
CAPÍTULO 2. MATERIALES Y MÉTODOS.....	12
2.1. Materiales.	12
2.2. Experimento.	13

2.2.1. Algoritmo de Hermes (SHS).....	13
2.2.2. SIFT.	17
2.2.3. Método cepstral (CEPSTRUM).....	18
2.2.4. Autocorrelación.	18
2.3. Modificaciones.	19
2.4. Relación del algoritmo de Hermes con el modelo auditivo.	20
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN.....	22
3.1. Contornos.	22
3.2. Resultados generales.....	30
CONCLUSIONES Y RECOMENDACIONES	33
REFERENCIAS BIBLIOGRÁFICAS.....	35
ANEXOS.....	38
ANEXO A Tabla de resultados de los errores por grupos disártricos.....	38
ANEXO B Listado del programa implementado en MATLAB ver 7.0.....	39

INTRODUCCIÓN

La comunicación humana es un factor clave para el desarrollo social de las personas y sobre todo su vía natural que es el habla. Los estudios de la voz han tenido gran interés en el ámbito internacional. Durante muchos años se han realizado una variedad de estudios de la señal de voz que han desarrollado tecnologías para la transmisión de dicha señal a grandes distancias (Hernández-Díaz, 1995).

También en el campo de la medicina se ha vinculado el procesamiento del habla para el diagnóstico y tratamientos de enfermedades y de esa manera contribuir a la calidad de vida de los pacientes. Las investigaciones tienen una tendencia al análisis de fonemas individuales y ahí se han reportado una mayor cantidad de trabajos que en el campo de la articulación y la prosodia (Ferrer, 2005) que analizan unidades más complejas del habla como palabras y oraciones.

Un parámetro que reviste gran importancia en el análisis de la voz lo constituye la detección de la frecuencia fundamental de la voz (F_0) que presenta una serie de dificultades para su cálculo. Se han desarrollado muchos algoritmos para (Hess, 1983) resolver este problema pero no han encontrado una solución óptima (Hernández-Díaz, 1995) que sea aplicable a todas las necesidades en general.

Con las variaciones de la periodicidad es posible obtener la entonación de los hablantes, aspecto este importante en la modelación prosódica. Los análisis de entonación se han realizado de manera subjetiva por los especialistas en lingüística, pero pueden existir patrones que no son perceptibles por el oído

humano y que se puede compensar a través de mediciones acústicas con ayuda de los medios de cómputo y las técnicas de procesamiento digital de señales.

En Laboratorio de Procesamiento de Voz del CEETI se han desarrollado algoritmos para la detección del período fundamental, orientados al cálculo del período pulso a pulso para mediciones como el *ruido*, *jitter* y *shimmer*, asociados a problemas de la generación de la vocalización. En el presente trabajo se desea medir la entonación en el habla fluida. La aplicación de los algoritmos implementados a voces patológicas incrementa el grado de complejidad en el análisis e interpretación de los resultados.

Las alteraciones en los patrones de entonación en muchos casos pueden tener valor diagnóstico. Existe una necesidad de encontrar un algoritmo para la detección de periodicidad que sea robusto a la hora de medir la entonación lo más precisa posible en personas que presentan deficiencias en el habla. Se debe seleccionar e implementar un algoritmo para la resolución de nuestro problema. De aquí se formula la siguiente interrogante:

¿Qué algoritmo de detección de la frecuencia fundamental de la voz puede ser útil para estimar de manera más eficiente la entonación en pacientes con trastornos en el habla?

Partiendo de la hipótesis de que el algoritmo de Hermes, por basarse en el modelo auditivo, debe determinar mejor el contorno de entonación que otros algoritmos que se basan en el modelo de producción del habla, y analizando el problema de nuestra investigación se derivan las siguientes interrogantes científicas:

¿Qué algoritmos detectores del período fundamental son más utilizados en la literatura?

¿Cómo se comporta el algoritmo de Hermes frente a otros algoritmos de detección de la frecuencia fundamental?

¿Cómo evaluar la efectividad de los algoritmos programados?

Para dar cumplimiento a la hipótesis planteada se propuso como objetivo general de la presente investigación:

Evaluar el modelo auditivo para la detección de la frecuencia fundamental en voces patológicas frente a otros algoritmos que se basan en el modelo de producción del habla.

De aquí se derivan los objetivos específicos que se muestran a continuación

- Seleccionar distintos algoritmos para la detección de F0.
- Programar los algoritmos con ayuda del software MATLAB ver. 7.0.
- Realizar un estudio comparativo para entre los diferentes algoritmos en cuanto a la cantidad de fallas en el cálculo de F0.
- Evaluar los resultados para determinar el más eficiente para hallar los contornos de entonación en voces patológicas.

Con este proyecto se pretende la selección del algoritmo más robusto para la modelación prosódica, fundamentalmente la detección de entonación. Los resultados derivados de esta investigación serán utilizados posteriormente en diferentes investigaciones que se realicen en el Laboratorio de Voz del CEETI y tiene aplicabilidad en la consulta de Logopedia y Foniatría.

El trabajo se estructura en introducción, capitulo, conclusiones, referencias bibliográficas y anexos.

En la introducción se define la importancia y necesidad de temática tratada.

El Capítulo 1 es el fruto de una extensa revisión bibliográfica. En el mismo se exponen las mediciones de la voz que se realizan y los problemas existentes en ellas y así como la propuesta de solución.

El Capítulo 2 trata los algoritmos seleccionados para la implementación y la vía para la validación del experimento.

El Capítulo 3 muestra los resultados validados a través de la comparación de los detectores implementados.

CAPÍTULO 1. MEDIDAS DE ENTONACIÓN

En este capítulo se realiza un análisis de las mediciones que se realizan en voces patológicas, así de los problemas que existen y la necesidad de correlacionar el análisis subjetivo con los parámetros objetivos en el diagnóstico médico. Se realiza una revisión bibliográfica para el estudio de los diferentes aspectos de la prosodia. Además se hace un estudio de diferentes afecciones que se reflejan en el habla humana y de cómo incluir el modelo auditivo en la implementación de un algoritmo que sea capaz de representar de la manera más precisa el contorno de entonación de pacientes con disartrias.

1.1. Problemas de las mediciones en voces patológicas.

Las mediciones en la señal de voz son de gran interés para el estudio de afecciones reflejadas en el habla humana por la importancia que encierra la comunicación oral en el contexto social de las personas. El procesamiento digital de dicha señal ha jugado un papel fundamental, pues permite realizar valoraciones objetivas a la hora del diagnóstico médico.

Existen mediciones tanto subjetivas como objetivas, en las cuales muchos han sido los trabajos realizados para encontrar el comportamiento de los patrones del habla (Kent & Ball, 2000). En las mediciones subjetivas el instrumento empleado es el oído humano mientras que entre las mediciones objetivas podemos encontrar las mediciones acústicas. En (Baken, 1999) se plantea que el oído del especialista es imprescindible pero no suficiente para un análisis integral de la voz del paciente. Es necesario vincular el análisis acústico al análisis subjetivo, pues se pueden presentar perturbaciones en la voz que el oído de un experto no es

capaz de detectar; por otra parte si al análisis acústico se le agrega una opinión especializada daría una valoración más integral de los problemas encontrados.

A través de los medios de cómputo tales análisis se pueden documentar de manera objetiva, además de la evolución de los tratamientos propuestos. La adquisición de las señales de voz se hace más fácil, por lo que la evolución del paciente tiene un seguimiento más confiable. El costo de estos permite la incursión de los mismos en las consultas médicas especializadas en el tratamiento del habla.

En la literatura se encuentran una gran cantidad de trabajos que emplean las mediciones relacionadas con el periodo fundamental para el estudio de la calidad vocal (Ferrer, 2005), como determinar periodicidad en vocales sostenidas. A medida que se incrementa la complejidad en las unidades del lenguaje se dificulta aún más la obtención de medidas objetivas para determinar los rasgos distintivos en el habla de los pacientes; esta es la causa fundamental de que los estudios se encaminen más a la parte de calidad vocal y no así en articulación y prosodia. Cada vez se hace más necesario desarrollar algoritmos que permitan encontrar patrones de periodicidad más allá de un fonema, como lo son en frases, oraciones y párrafos. Los patrones en el habla fluida van a dar más precisión a la hora de diagnosticar enfermedades que se reflejan en el habla de las personas y según sea su comportamiento y profundidad se puede dar el tratamiento adecuado.

Todas estas mediciones en conjunto, junto con las realizadas en pacientes que presentan nasalidad interactúan entre si, por ejemplo, pacientes con problemas prosódicos pueden tener en conjunto afectaciones en su articulación, problemas de nasalidad. Con una aproximación general de todos estos patrones encontrados en las diferentes mediciones se puede realizar una buena aproximación para el estudio de la inteligibilidad.

1.2. Afecciones en el habla.

En la producción del habla intervienen un gran número de sistemas y órganos y cualquier problema en ellos puede desencadenar algún problema en la misma.

Entre las afecciones podemos mencionar las apraxias y las afasias que son desórdenes neurológicos que afectan la información lingüística.

Las disartrias son desórdenes complejos en el habla, pues representan una variedad de afecciones neurológicas que pueden afectar potencialmente cada componente del aparato productor del habla. Estas pueden reflejarse por la presencia de tumores, infecciones, traumatismos, o causas desconocidas (Ferrer, 2005).

1.3. Análisis acústico en la caracterización de disartrias.

El análisis acústico es un valioso complemento a la evaluación perceptiva de un especialista (Kent & Weismer, 1999). Los estudios en el ámbito de las disartrias es un reto, pues estas son desórdenes complejos que ocurren a lo largo del sistema de producción del habla y que a su vez pueden enmascarar otros desórdenes. Los estudios acústicos ofrecen una valoración cuantitativa que se correlaciona con el juicio perceptual de inteligibilidad, calidad y tipo de disartria.

Aunque las valoraciones perceptuales se consideran las bases para la clasificación y descripción de las disartrias se han planteado preguntas sobre la validez y fiabilidad de las mismas, sobre todo cuando los juicios son emitidos por personas diferentes o son dados por especialistas que no tienen el entrenamiento suficiente para la práctica perceptual (Kent, 1996; Zyski & Weisiger, 1987). También se ha cuestionado si estos solo pueden usarse para diferenciar entre afectaciones que ocurren en dos o más componentes del sistema de producción del habla.

Las valoraciones acústicas han sido recomendadas para complementar las valoraciones perceptuales, pues estas superarán limitaciones que presentan los métodos subjetivos (Collins, 1984). Además puede ser valioso como una fuente objetiva de datos para las valoraciones clínicas.

Las disartrias traen consigo perturbaciones en los rasgos prosódicos, la búsqueda de un método eficaz para su análisis está aún bajo estudio. Uno de los problemas que se presentan es la selección del material en análisis, pues detectar problemas

prosódicos no es igual si se toma la lectura de un párrafo que el análisis a un fragmento de una conversación natural. En otros estudios se plantea que la conversación puede ser más útil en la detección de estas anomalías, pero tiene la desventaja que en ella se pierden propiedades del habla como la pronunciación y la composición sintáctica. **Prosodia.**

El término prosodia (Crystal, 1982) recoge una serie de aspectos de la producción del habla como son la entonación, la razón del habla, la unión, la acentuación y el ritmo, que su uso no solo se aplica a un fonema por separado, sino a unidades del habla con mayor grado de dificultad, como palabras, frases, oraciones, párrafos. Aunque también se debe tener en cuenta el efecto que causa un sonido por separado en una sílaba y las consecuencias que trae para las mediciones que se realizan. El conjunto de todos estos análisis se le conoce como análisis suprasegmental del habla (Lehiste, 1970; Shriberg & Kent, 1982). En la prosodia podemos encontrar diferentes afectaciones como igual acentuación y excesiva, ráfagas cortas del habla, silencios inapropiados y monotonías de intensidad y tono. **1.4.1. Entonación.**

La entonación es la variación sistemática del tono de la voz (Odhe & Sharf, 1992). El tono es la percepción de la frecuencia fundamental (F0), o sea, la frecuencia de vibración de las cuerdas vocales. La percepción de F0 es relativa, pues una misma frase puede ser pronunciada con intenciones diferentes por lo que la entonación dada a cada expresión también difiere. El valor de la frecuencia fundamental tiende a crecer cuando se realiza una pregunta, decrece si la expresión es una afirmación completa y se mantiene constante en el caso de una afirmación incompleta. Además estos cambios no son interpretados de la misma manera en voces de hombres, mujeres y niños, cuyos valores de F0 se encuentran entre 120 y 130 Hz, 190 y 225 Hz y mayores que 256 Hz respectivamente (Tiffany & Carrell, 1977).

Las variaciones relativas de F0 son representadas a través de un contorno que representa la entonación del hablante. El contorno de entonación representa una sílaba o la unión de sílabas acentuadas o no, donde a través de esta relación se

aprecia el cambio del patrón de F0 (Odhe & Sharf, 1992). En el análisis del habla fluida se identifican sílabas que tienen mayor peso que las otras y que son identificadas con valores de pitch altos acompañado por una duración larga y una mayor fuerza a la hora de la pronunciación. Los cambios en la entonación implican cambios en la acentuación y la duración de las sílabas.

La frecuencia fundamental se puede determinar de diferentes maneras. Tomando como punto de partida que la señal del habla no es estacionaria y varía con el tiempo se tienen que tomar diferentes estrategias, por ejemplo, se debe tomar un segmento corto de señal para realizar el análisis y el dominio en que se desea trabajar (tiempo, frecuencia, transformado, híbrido). Para la selección del algoritmo se debe tener en cuenta primeramente la aplicación que se le va a dar al mismo, así como la cantidad de errores que arroja en las mediciones y el costo computacional. En igualdad de condiciones los resultados que presentan diferentes PDAs (Pitch Detection Algorithm, algoritmos de detección del pitch) pueden variar.

1.4.2. Problemas en los contornos de entonación.

Existen dos tipos fundamentales de trastornos en la entonación de los hablantes, uno que afecta la cantidad de contornos de entonación y otro el cual afecta el grado de fluctuación del tono. Pocos contornos de F0 aparecen en el habla rápida de algunos pacientes afásicos receptivos o en el habla lenta en pacientes con retardo mental o con afasia expresiva.

Los patrones fundidos, los cuales aparecen en pacientes con Parkinson y disartria del hemisferio derecho, casi no muestran cambios en la frecuencia fundamental. Carencia de variaciones de F0 es evidencia de patrones disociados que son características de pacientes con disartria atáxica y algunos con apraxia. Sin embargo en los patrones disociados no hay continuidad en F0 de sílaba a sílaba, como ocurre en los patrones fundidos.

1.5. El modelo auditivo.

Un detector de frecuencia fundamental confiable es imprescindible en la modelación prosódica, sobre todo en la representación de los contornos de entonación. Las voces de los pacientes aquejados por disartrias presentan características peculiares, logrando así que algoritmos robustos presenten problemas en el cálculo de la frecuencia fundamental. Específicamente los errores que presentan los algoritmos radican principalmente en la decisión de sonoridad y debido a las características de la señal sonora en el propio cálculo de F0.

El modelo auditivo es muy utilizado en la implementación de algoritmos para la representación de rasgos de entonación en la música y no se conoce su desempeño en la modelación prosódica. Por otra parte, las medidas de entonación se han realizado por los especialistas en lingüística de manera subjetiva. Para correlacionar las opiniones de los expertos, con medidas acústicas, este trabajo explora el beneficio de introducir el modelo auditivo, en el algoritmo para la detección del periodo, así se contribuye a obtener un contorno más preciso de las variaciones de la frecuencia fundamental en los pacientes con disartrias (Shrivastav, 2003).

Muchos PDAs han tratado de realizar una representación del proceso de detección de tonos en el sistema auditivo humano (Hermes, 1988). En (Duifhuis et al., 1982) se realizó una implementación de la teoría de la percepción de Goldstein (Goldstein, 1973; Hernández-Díaz, 1996). Ellos consideraron esta teoría cuantitativa de la percepción subjetiva factible para aplicaciones prácticas.

También se desarrolló la teoría del pitch virtual propuesta en (Terhardt, 1979, 1982), que asume que cada componente espectral genera una serie de subarmónicos que se combinan y posibilita la percepción de la frecuencia fundamental. El concepto anterior fue explícitamente formulado por de Boer (de Boer, 1977). Esta idea incorpora la presencia de una serie de elementos sensibles a la frecuencia, que no solo se activan en la frecuencia fundamental que es donde presenta mayor sensibilidad, sino también aquellas frecuencias mayores

que tienen relación con la más sensible. Por ejemplo, si la frecuencia fundamental es de 200 Hz también se activan las componentes de frecuencias de 400, 600, 800 Hz, etc. En (Houtgast, 1976) se encontraron evidencias experimentales de que los subarmónicos son generados en el procesador central del período.

Además se implementaron dos algoritmos para el análisis del habla. Los mismos se basan en el principio de que la frecuencia fundamental tiene sus raíces en la generación y combinación de subarmónicos. El primero es el algoritmo de compresión espectral (Schroeder, 1968) que consiste en la suma o multiplicación de una secuencia de espectros comprimidos armónicamente. El otro es el algoritmo implementado por (Martin, 1982), donde se realiza un procedimiento equivalente a una sumatoria del espectro comprimido, donde la distancia entre los picos se reduce cada vez que el espectro se comprime. Según (Hess, 1983) los algoritmos basados en la compresión espectral son los más confiables por corresponderse mejor con las teorías del pitch.

1.6. Selección del algoritmo.

Para la modelación prosódica, sobre todo para la representación de los contornos de entonación en voces patológicas se implementa el algoritmo de SHS publicado por (Hermes, 1988) y es una correspondencia con la teoría perceptual para la detección de periodicidad. El mismo utiliza muchos conceptos planteados hasta ese momento y e incorpora otros con los cuales se gana eficiencia en los cálculos para determinar F0. Se calcula una transformada de Fourier de 256 puntos y se realiza una interpolación a 48 puntos por octava. El espectro se interpola en un eje logarítmico, pues así la distancia entre los picos permanece constante al igual que la resolución de los mismos; además como una compresión en un eje lineal se corresponde con un desplazamiento en el eje logarítmico es beneficioso a la hora de realizar la sumatoria.

El algoritmo de Hermes es muy robusto cuando hay presencia de ruido en la señal y este trabajo se encarga de demostrar su robustez para la representación de la entonación de los pacientes con disartrias. Al trabajar en el dominio de la

frecuencia debe resolver los problemas en la detección de periodicidad en voces patológicas, pues debe arrojar mejores resultados que otros algoritmos basados en el modelo de producción del habla.

El estudio de la implementación del algoritmo de Hermes va a estar encaminado a encontrar los contornos de entonación en pacientes que padecen de disartrias. Lo anterior no quiere decir que no sea aplicable en otros campos. En pacientes sordos, tartamudos o que presentan otras anomalías se puede aplicar el modelo auditivo tanto en el análisis como en los tratamientos de sus problemas.

Conclusiones parciales

A través del estudio de este capítulo queda expuesta la necesidad de insertar el modelo auditivo en los PDAs para el análisis clínico, se ha escogido el algoritmo de Hermes, al tener una estrecha relación con este modelo, se explora su capacidad para representar los contornos de entonación de manera más precisa que otros que se basan en otros modelos de producción del habla.

CAPÍTULO 2. MATERIALES Y MÉTODOS

En el transcurso de este capítulo se realiza una descripción de los algoritmos implementados y de los parámetros a tener en cuenta para la detección de la frecuencia fundamental aplicado a voces patológicas, además de las herramientas utilizadas en el proceso. Se definen los materiales con los que se cuenta y el diseño experimental para la validación de los datos.

2.1. Materiales.

Para realizar la detección de los contornos de entonación se toman grabaciones de pacientes de la base de datos de Aronson (Aronson, 1993), que cuenta con una muestra de personas que presentan diferentes tipos de disartrias. De ella se selecciona un grupo de 18 pacientes (12 mujeres y 6 hombres) que pertenecen a seis grupos disártricos diferentes. Cada uno de los seleccionados pronuncia tres oraciones en idioma inglés.

Las grabaciones de las voces se convierten a un formato PCM, con frecuencia de muestreo de 11.025 kHz y con resolución de 16 bits con un canal (mono) a razón de 21 kb/s. El almacenamiento se efectúa en un archivo con extensión .WAV utilizando para todo el proceso la herramienta "Sound Recorder" que brinda el sistema operativo Windows XP.

Los algoritmos detectores de periodicidad seleccionados son implementados con la ayuda del software MATLAB, versión 7.0, por las bondades que brinda en el cálculo, programación y visualización en pantalla de los datos de interés. Además el Toolbox de Procesamiento Digital de Señales posee una serie de funciones útiles para el análisis de las señales de voz e imágenes.

Con la utilización del software ECAH, desarrollado en el CEETI, se pueden realizar diferentes análisis de la señal de voz. El mismo brinda la posibilidad de observar el espectrograma de banda estrecha, que no es más que la estructura de los formantes de la voz de los pacientes y que es útil para la validación del experimento. También juega su papel como herramienta de edición pues permite seleccionar determinados segmentos de la señal que presentan niveles de ruido muy altos y pueden ser suprimidos para que no afecten las mediciones que se realizan.

2.2. Experimento.

Para realizar una valoración del desempeño del algoritmo de Hermes en la detección de periodicidad en voces patológicas se implementan otros tres PDAs que se basan en el modelo de producción del habla para realizar una comparación entre ellos y seleccionar el más adecuado para dichas mediciones. Se toma como referencia el espectrograma de banda estrecha del ECAH para determinar los intervalos sonoros y de silencio en las voces, y así realizar un conteo de las tramas que arrojaron estimaciones erróneas en el cálculo de F0. Los resultados son expuestos en el capítulo 3. **Algoritmo de Hermes (SHS).**

Para la valoración del funcionamiento del modelo auditivo en la detección de periodicidad en voces patológicas se implementa el algoritmo de Subharmonic Summation (SHS) propuesto por (Hermes, 1988). El favorable desempeño del algoritmo radica en que se corresponde con la teoría de percepción del pitch.

El mismo es un detector de periodicidad que trabaja en el dominio de la frecuencia y se sustenta sobre la base de que cada componente espectral genera una serie de subarmónicos que se combinan y dan lugar a la percepción del pitch, o lo que es equivalente a que cada componente espectral activa no solo esos elementos del procesador central del oído que son más sensibles a la frecuencia de esa componente, sino también a esos elementos que tienen una relación de armónicos menor con la misma. Estos elementos se adicionan y la activación será mayor en el elemento de frecuencia más sensible dando como resultado la estimación de F0. Este método pone en práctica que una compresión en la escala lineal de

frecuencia se corresponde con un desplazamiento en un eje logarítmico. Con tales desplazamientos la distancia entre los picos se mantiene constante, entonces solo es necesario determinar la contribución de los valores de frecuencia a la detección de la periodicidad. La resolución de los picos tampoco varía con dichos saltos.

La figura 2.1 muestra el diagrama en bloques propuesto para la implementación del detector de F0. El mismo cuenta con siete pasos que da como resultado la detección de la periodicidad de la señal de voz.

1. La señal de voz que se obtiene al realizar la grabación de los pacientes se enventana con el propósito de obtener segmentos de 40 ms de duración con frecuencia de muestreo igual a 10 kHz. Se le realiza un downsampling 4:1.
2. Dicha señal se multiplica por una ventana de Hamming y posteriormente se le aplica una FFT de 256 puntos que brinda una buena resolución para los análisis posteriores.
3. Al espectro se le enfatizan los picos según (Martin, 1982). El espectro es suavizado a través de un filtro de Hanning.
4. Se realiza el cambio de escala de los ejes de frecuencia, de lineal a logarítmica. Los valores del espectro se obtienen a través de una interpolación cubic-spline a 48 puntos por octava de resolución.
5. El espectro obtenido en el paso anterior se multiplica por una ventana \tan^{-1} elevada, la cual representa la respuesta de frecuencia del oído humano para frecuencias por debajo de 1250 Hz.

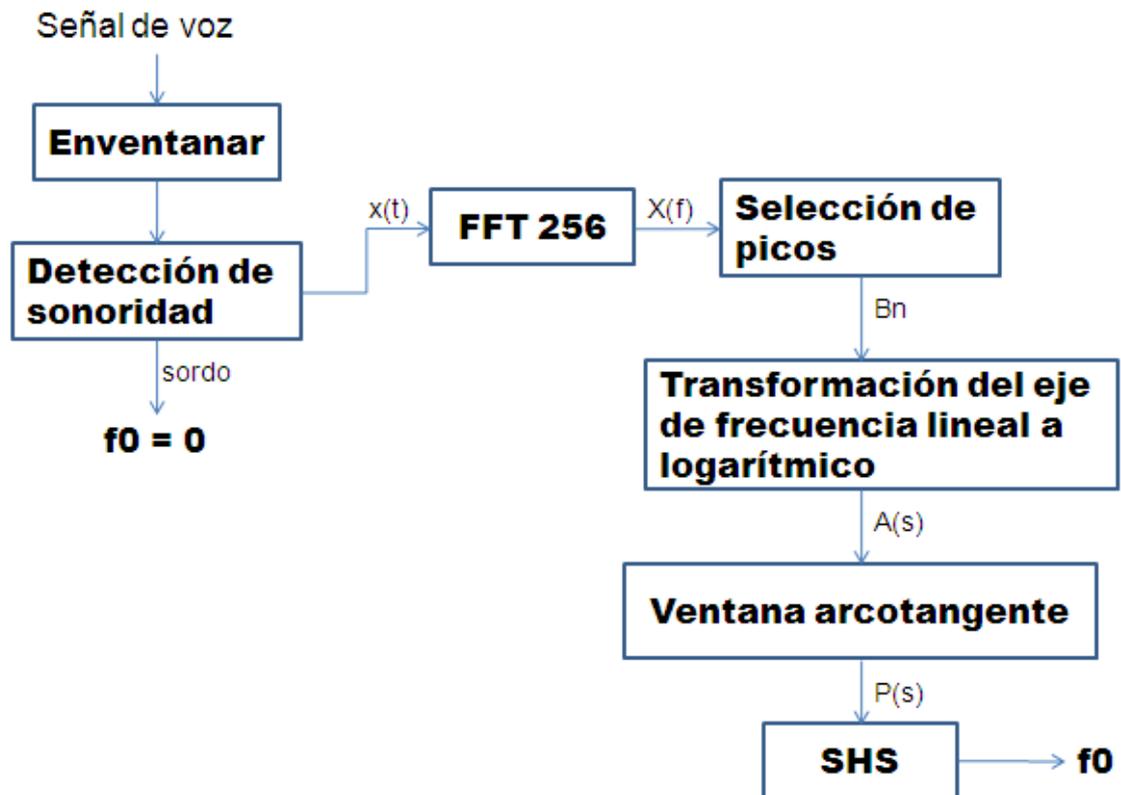


Fig. 2.1. Algoritmo de Hermes.

6. Se realiza la adición de los subarmónicos desplazando el espectro en el eje logarítmico (compresión en el eje lineal) de frecuencia multiplicándose por un factor.

$$H(s) = \sum_{n=1}^N h_n P(s + \log_2 n) \quad (1)$$

ó en la escala lineal:

$$H(f) = \sum_{n=1}^N h_n P(nf); \quad (2)$$

donde:

$$s = \log_2 f \quad (3)$$

N es la cantidad de subarmónicos que se tienen en cuenta a la hora del cálculo de la frecuencia fundamental, **n** representa el factor de compresión

del espectro y h_n refleja que los armónicos en altas frecuencias tienen menos peso en la detección de F_0 , este último se calcula como:

$$h_n = 0.84^{n-1} \quad (4)$$

El valor de la frecuencia fundamental lo constituye el valor máximo de $H(s)$, este último representa el espectro de la sumatoria de los subarmónicos.

En la figura 2.2 se representa un ejemplo de la sumatoria de subarmónicos donde se tienen en cuenta un valor de $N=5$ a la hora de calcular el valor de F_0 .

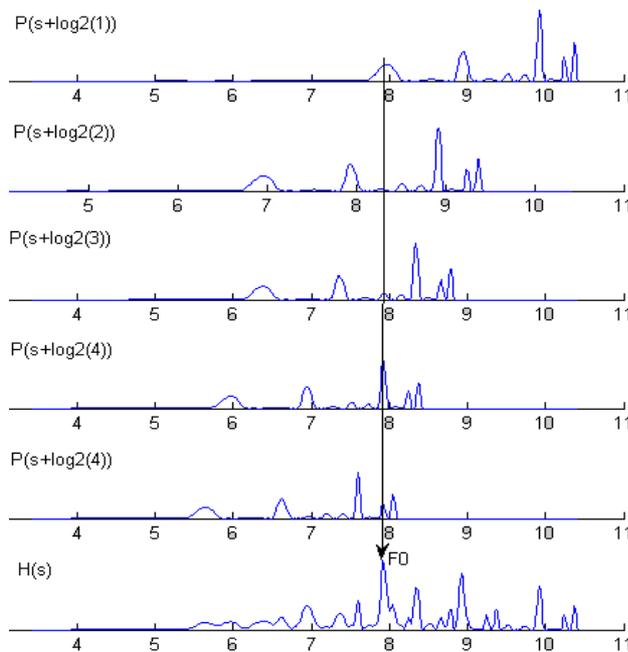


Fig. 2.2. SHS para tomando 5 subarmónicos.

7. La detección de sonoridad está dada por un valor de autocorrelación de la trama en análisis. Se considera la trama sonora cuando el valor sobrepasa el umbral de 0.52.

Para la implementación del algoritmo de Hermes la señal de voz se divide en segmentos de 40 ms de duración y la separación entre cada valor de F_0 detectado existe una separación de 10 ms. La interpolación se realiza a 48 puntos por

octava pues brinda buena resolución en el eje logarítmico de frecuencia para la detección de periodicidad.

2.2.2. SIFT.

El algoritmo basado en la técnica de filtrado inverso simplificado (SIFT) (Markel, 1972) trabaja en un dominio híbrido, pues emplea el análisis LPC (Lineal Prediction Coeficients, coeficientes de predicción lineal) mencionadas en (Hernández-Díaz, 1997) y el análisis de autocorrelación. En la figura 2.3 se muestra el diagrama en bloques utilizado para su implementación. A diferencia del algoritmo original, este careció de un diezmo porque los medios de cómputo utilizados soportan un volumen grande de datos para el procesamiento; además de una interpolación después de la función de autocorrelación pues esto compensaba lo anteriormente dicho. Se utiliza una frecuencia de muestreo de 11.025 kHz y tramas de 40 ms de duración.

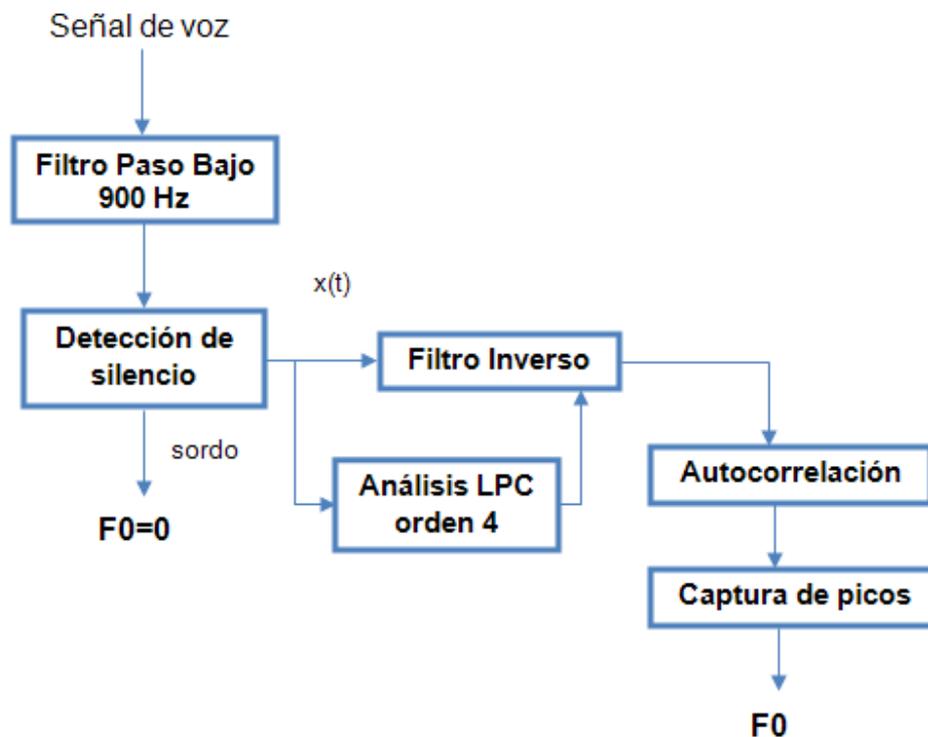


Fig. 2. 3. Diagrama en bloques del SIFT.

2.2.3. Método cepstral (CEPSTRUM).

El algoritmo basado en el método cepstral para la detección de periodicidad basa su implementación en el cálculo de la transformada inversa de Fourier del logaritmo del espectro de potencia de la señal por lo que trabaja en un dominio transformado. Se utiliza el término cepstrum para la expresión 5 que involucra solo el logaritmo de la magnitud del espectro.

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|X(e^{j\omega})| e^{j\omega n} d\omega \quad (5)$$

En la figura 2.4 se muestra el diagrama en bloques de dicho algoritmo (Noll, 1967).

Para la implementación del método cepstral se cambió el análisis de cruces por cero por una detección de sonoridad basada en la energía de la señal de voz de 40 ms de duración cada una.

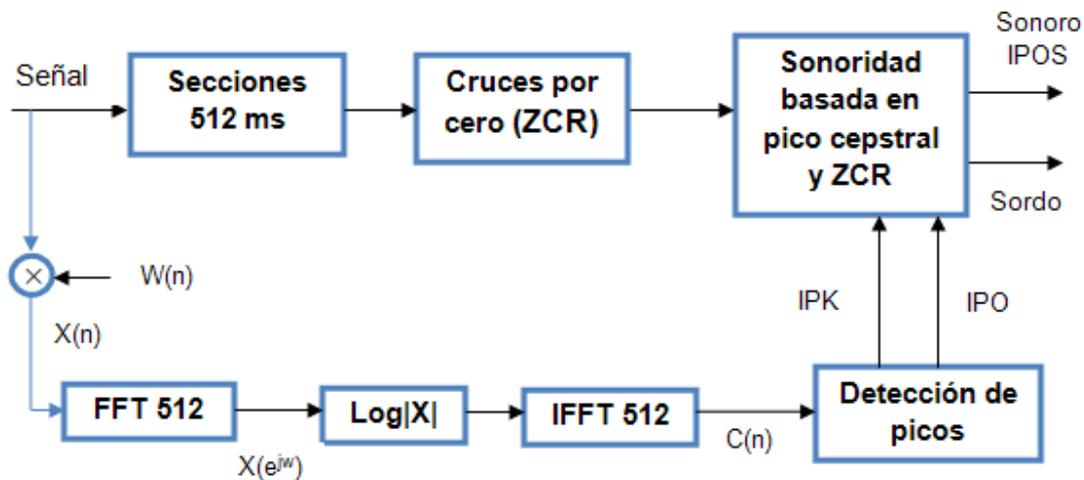


Fig. 2. 4. Diagrama en bloques del método cepstral.

2.2.4. Autocorrelación.

En el análisis de periodicidad se implementó el algoritmo de autocorrelación propuesto por (Dubnowski, 1976), que se basa en el análisis de la señal temporal. En la figura 2.5 muestra el diagrama en bloques de la implementación que se diferencia de la original por carecer del cálculo de niveles de recortamiento central, manteniendo solo un proceso de detección de sonoridad.

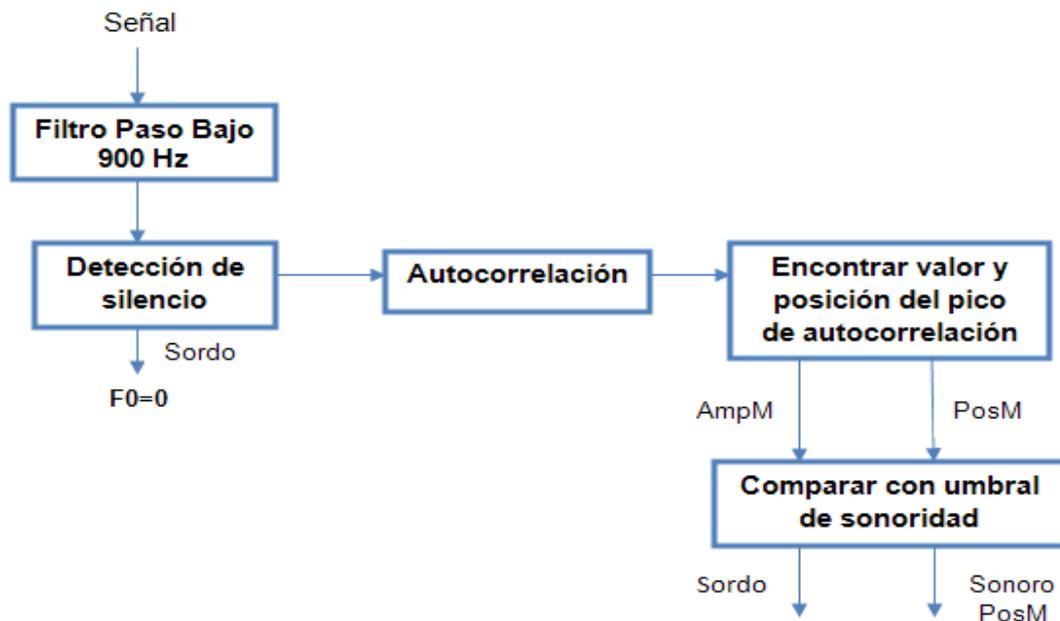


Fig. 2.5. Diagrama en bloques del algoritmo de Autocorrelación.

2.3. Modificaciones.

El algoritmo de SHS para obtener los contornos de entonación en voces patológicas sufrió una serie de cambios debido a las características de las grabaciones y de las voces de los pacientes.

La frecuencia de muestreo utilizada es de 11.025 kHz, esto implica que en cada trama en análisis la cantidad de muestras presentes es mayor y se considera que frecuencias por encima de 1370 Hz no son tomadas en cuenta en la detección del pitch. El espectro después de aplicar la FFT va a consistir en valores de amplitud de 128 puntos espaciados 10.7 Hz en el eje de frecuencia lineal. El procesamiento posterior no afecta en gran medida la precisión del cálculo de interés. La sumatoria de subarmónicos siempre va a tener un máximo en cualquier lugar del espectro por lo que se definen umbrales en los valores de F0, un mínimo de 50 Hz y a su vez no se toman en cuenta valores mayores que 1250 Hz; este último es muy poco probable en voces afectadas por lo que no se le considera en la programación del algoritmo.

La detección de tramas sonoras y sordas se realiza a través de una lógica de umbrales de autocorrelación y valores de energía de cada fragmento en análisis. Para la autocorrelación se toma un valor de 0.4 por la falta de periodicidad en la señal de voz de los pacientes y los rápidos cambios que pueden afectar la medición. Además se considera sonora la trama que su nivel de energía sobrepase por lo menos cinco veces el valor de energía mínimo de la señal.

A cada contorno detectado se le realiza un suavizamiento de 5 puntos de mediana (Hess, 1983) que ayuda además a una mejor definición de los segmentos sordos y sonoros. Esto último y la detección de sonoridad se le aplican también al resto de los algoritmos. En el capítulo 3 se reflejan los resultados del desempeño de los PDAs tanto sin suavizamiento que suavizados.

En el anexo B se presenta la implementación del algoritmo de Hermes.

2.4. Relación del algoritmo de Hermes con el modelo auditivo.

El algoritmo SHS es una aplicación confiable del concepto de que los subarmónicos son realmente generados en el procesador central del cerebro (de Boer, 1977). El hecho de que los armónicos mayores tienen menor contribución a la detección de F_0 (Houtsma & Goldstein, 1972) se representa por los valores decrecientes de h_n provocado por los aumentos de n . La sumatoria de los subarmónicos $H(s)$ puede verse como una aproximación de la activación combinada de todos los armónicos.

El algoritmo de Hermes presenta una aplicación continua de la teoría del pitch virtual (Terhardt, 1979) de que los subarmónicos son producto de todas las componentes espectrales. El algoritmo de Terhardt (Terhardt, 1982) parte de una lista de frecuencias y amplitudes detectadas. Para la extracción del pitch virtual se plantean cuatro principios y el peso numérico que se le atribuye al significado del cálculo del mismo. Se muestra que aunque el SHS no trabaja con un conjunto de picos detectados, opera de acuerdo con estos principios.

El primer punto planteado por Terhardt es el número de componentes espectrales relevantes que contribuyen al mismo pitch virtual; esto se representa esencialmente por la sumatoria en SHS. Segundo, el peso del pitch virtual debe

incrementarse con el peso del pitch espectral; en SHS se comprende como que cada componente espectral contribuye a la detección de F_0 con su amplitud multiplicada por la respuesta de frecuencia del oído humano. El tercer principio se basa en que dichos pesos deben disminuir con el número de armónicos, correspondiéndose con los valores decrecientes de h_n . Por último, si una componente difiere de la estructura armónica de los otros componentes, su peso debe disminuir en el cálculo de la periodicidad; en el algoritmo de Terhardt los decrementos son lineales, mientras que en el de Hermes se logra mediante el procedimiento de enfatizar los picos.

Una diferencia conceptual entre la teoría de Terhardt y la de Hermes es que SHS no distingue entre los pitch espectrales que están presentes físicamente en la señal de los virtuales que se levantan por la suma de los subarmónicos. En SHS el valor máximo de la sumatoria es el que se considera para la estimación de la frecuencia fundamental, pero los picos secundarios también son notables en otras frecuencias, especialmente las mayores componentes espectrales. En SHS no hay diferencias esenciales entre esos picos y en sus frecuencias, un valor de F_0 puede percibirse posiblemente, aunque los timbres sean diferentes. Estas diferencias en los timbres son explicadas asumiendo que el procesador central guarda registros de los subarmónicos que dan lugar al pitch. Escuchando analíticamente puede ser identificado con concentrarse en el sonido con el timbre de una onda sinusoidal pura.

Conclusiones parciales

El algoritmo de Hermes presenta buenas características para detectar las variaciones de la frecuencia fundamental de la voz, pues representa muy cercanamente el proceso de escucha que tiene lugar en el oído humano. El algoritmo original está diseñado para voces normales y es muy robusto para la modelación del habla telefónica, por eso es necesario variar los parámetros para lograr los resultados esperados del mismo aplicados a voces patológicas.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En este capítulo se muestra el desempeño de los algoritmos implementados y se presentan varias gráficas que reflejan los contornos detectados por cada uno en el contexto de una oración. Para los análisis se toman las oraciones de los pacientes pronunciadas en el idioma inglés. Al final se hace el comentario de los resultados globales que se obtienen con todas las tramas analizadas y se presentan en tablas.

3.1. Contornos.

A continuación se muestran algunos de los resultados obtenidos en las pruebas de los algoritmos programados para el habla fluida. Se escoge un ejemplo de cada oración para detectar sus contornos de entonación que vendrán acompañados de sus respectivos espectrogramas. De esta manera se muestran algunas de las características del habla de los pacientes disártricos y las diferencias entre las mediciones que cada detector realiza de la misma señal de voz.

En la figura 3.1 se muestra el espectrograma de banda estrecha de una mujer con la enfermedad de Parkinson al pronunciar la oración *Do you wish to know all about my grandfather?*, y en las figuras 3.2 y 3.3 se representan los contornos de entonación sin suavizar y suavizado respectivamente.

Se observa en el espectrograma que la paciente presenta una estructura de armónicos definida aunque por encima ya de 1500 Hz presenta muy pocos formantes; también se definen bien las zonas de silencio y donde existe sonoridad en la oración. Producto a su enfermedad presenta una voz monótona, es decir,

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

que no varía su frecuencia fundamental en toda la oración, lo que es bien observado en la gráfica de los contornos. Además no se le da la entonación necesaria de acuerdo a la idea que se quiere transmitir porque la oración es una pregunta y estos contornos se caracterizan por tener una tendencia a aumentar la frecuencia fundamental, sobre todo al concluir la oración propiamente dicha, sin embargo parece que estamos en presencia de una afirmación inconclusa. Esto se refleja claramente en las mediciones efectuadas por los cuatro algoritmos.

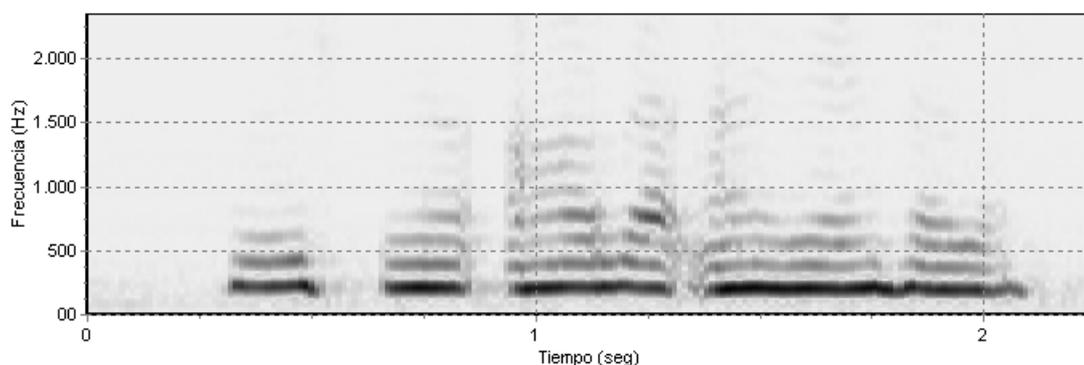


Fig.3.1. Espectrograma de la oración *“Do you wish to know all about my grandfather?”*

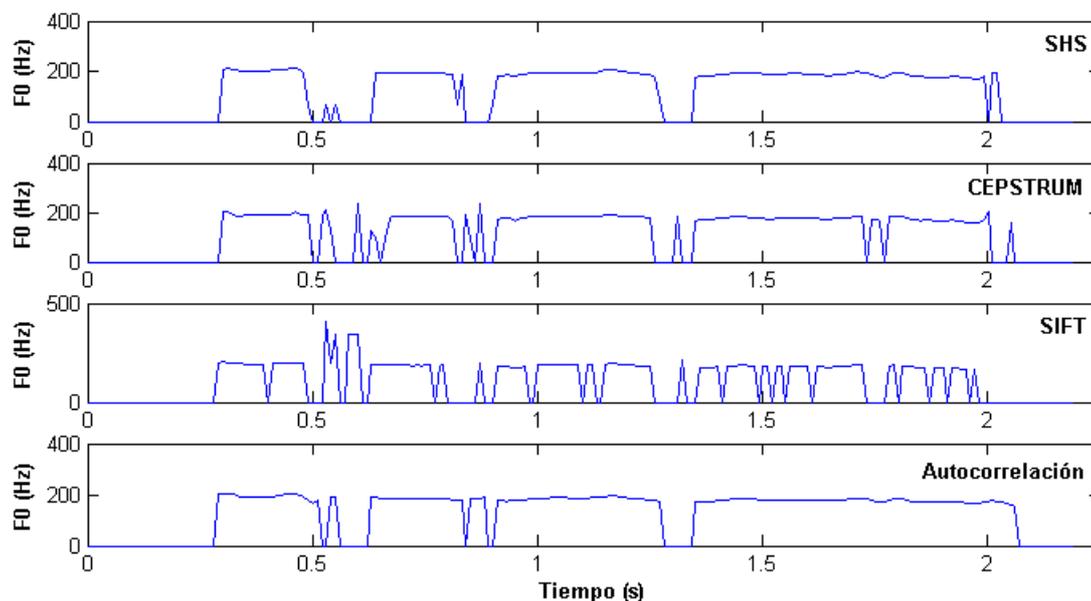


Fig.3.2. Contornos sin suavizamiento de *“Do you wish to know all about my grandfather?”*

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

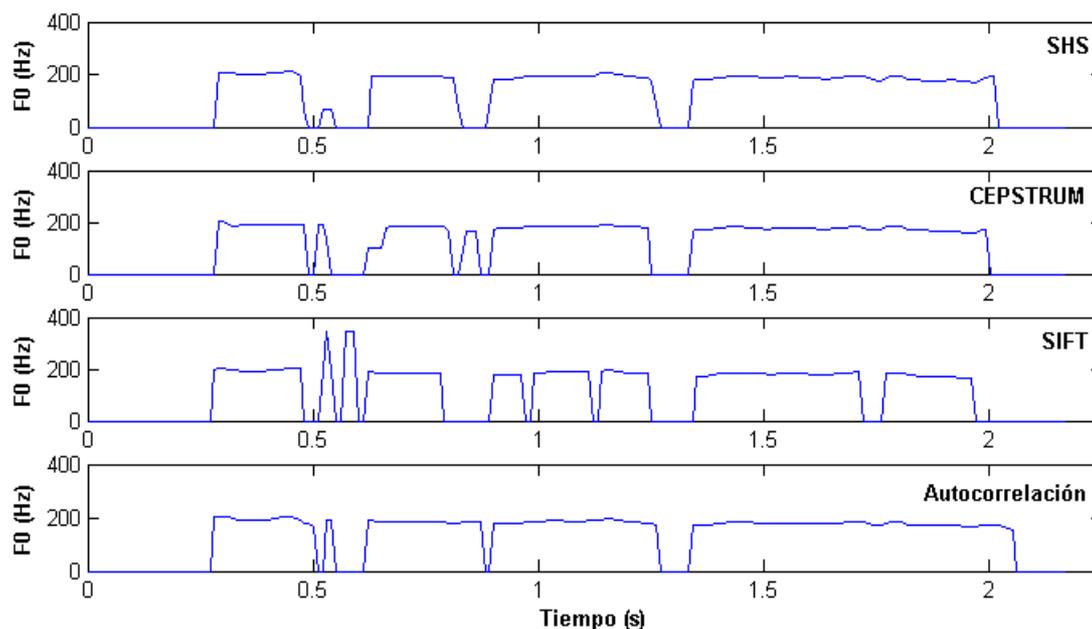


Fig.3. 3. Contornos suavizados de “Do you wish to know all about my grandfather?”

Algoritmos	EDT Sonoras	EDT Sordas
SHS	1 (0.45%)	2 (0.91%)
AUTOCORRELACIÓN	1 (0.45%)	2 (0.91%)
CEPSTRUM	5 (2.27%)	7 (3.18%)
SIFT	12 (5.45%)	20 (9.1%)

Tabla I. Errores de los contornos sin suavizamiento de la oración “Do you wish to know all about my grandfather?”

Algoritmos	EDT Sonoras	EDT Sordas
SHS	0 (0%)	2 (0.91%)
AUTOCORRELACIÓN	0 (0%)	2 (0.91%)
CEPSTRUM	2 (0.91%)	3 (1.36%)
SIFT	10 (4.5%)	6 (2.73%)

Tabla II. Errores de los contornos suavizados de la oración “Do you wish to know all about my grandfather?”

El SHS y el algoritmo de Autocorrelación representan mejor el contorno que los demás, como lo demuestran las tablas I y II, cuyas columnas muestran los errores en la detección de tramas (EDT) sonoras y en las tramas sordas. El principal error y que fue común en todos los PDAs fue el cambio de tramas sordas a sonoras

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

entre el primer y el segundo bloque continuo. En el espectrograma el color más oscuro, que es continuo constituyendo una línea a lo largo del eje de tiempo, representa los lugares donde hay presencia de sonoridad. A partir de 0.5 segundos existe una zona de silencio que no son bien detectadas por los algoritmos, los cuales reflejan valores de frecuencia fundamental que no deben estar presentes. También se observan rupturas en los bloques en el CEPSTRUM y el SIFT y en este último se pueden ver varios errores gruesos cometidos en la medición. Los errores gruesos son fallas drásticas del algoritmo detector del período (Rabiner et al, 1976). Las rupturas son debido a cambios de tramas sonoras en sordas muchas de las cuales son rectificadas mediante el suavizamiento del contorno.

En las figuras 3.4, 3.5 y 3.6 se muestran el espectrograma y los contornos de entonación sin suavizamiento y suavizados respectivamente, los cuales fueron detectados por los algoritmos implementados. Se presenta a una mujer con disartria atáxica pronunciando la oración *Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever.*

Esta paciente presenta una de las características que pueden aparecer con la presencia de su padecimiento que son pausas largas en lugares inapropiados. En el espectrograma se aprecian una serie de pequeños bloques que son el reflejo del habla lenta a partir de los cuatro segundos. Además se observa de manera clara los intervalos de tiempo donde se presentan los silencios en la oración.

Los contornos reflejan cierta monotonía en la entonación del hablante, aunque no tan pronunciada como en el caso anterior. Se muestran rupturas en los bloques en todos los algoritmos, fundamentalmente en el SIFT.

Los resultados que arrojó el análisis de los contornos de entonación se presentan en las tablas III y IV. El CEPSTRUM y el SHS muestran los mejores contornos según la cantidad de tramas erróneas. Las principales dificultades se presentaron en errores gruesos presentes en el de Autocorrelación y en el SIFT. También ocurrieron cambios de sonidos sonoros a sordos, mientras que a la inversa los detectores fallaron poco.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

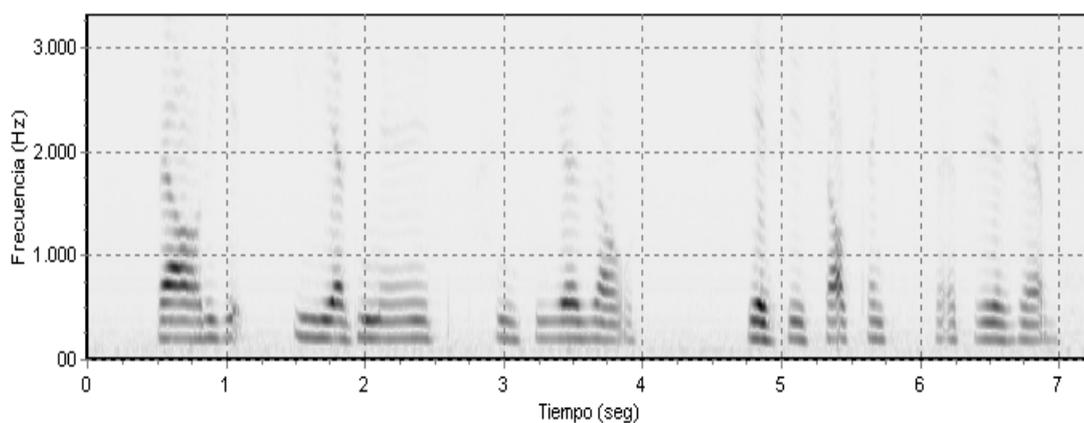


Fig.3.4. Espectrograma de la oración “*Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever*”.

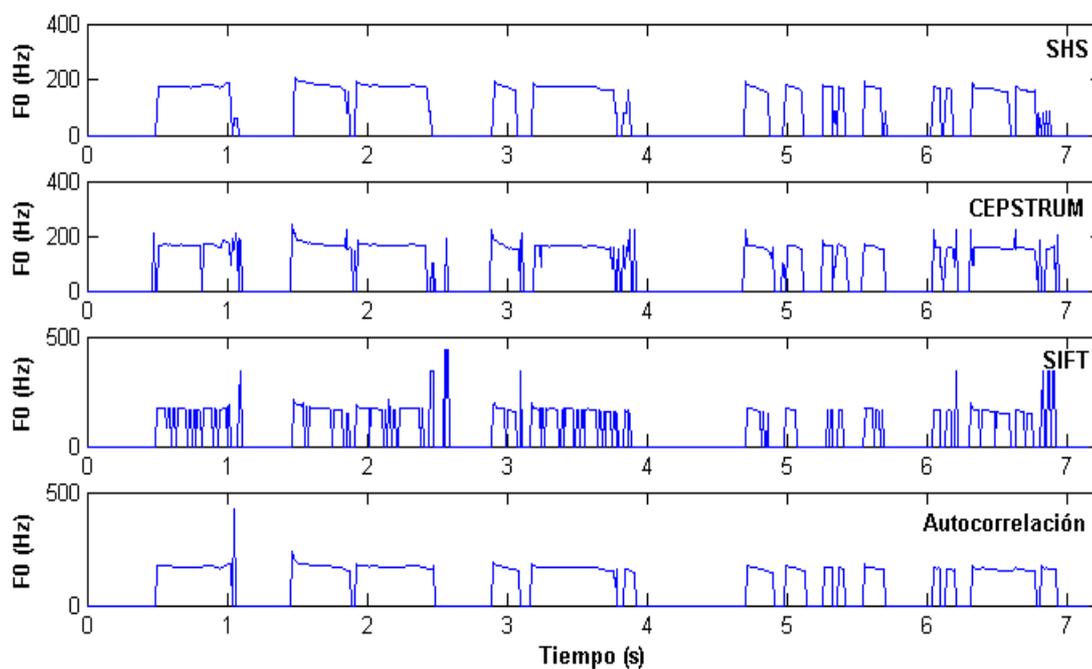


Fig.3.5. Contornos sin suavizamiento de “*Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever*”.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

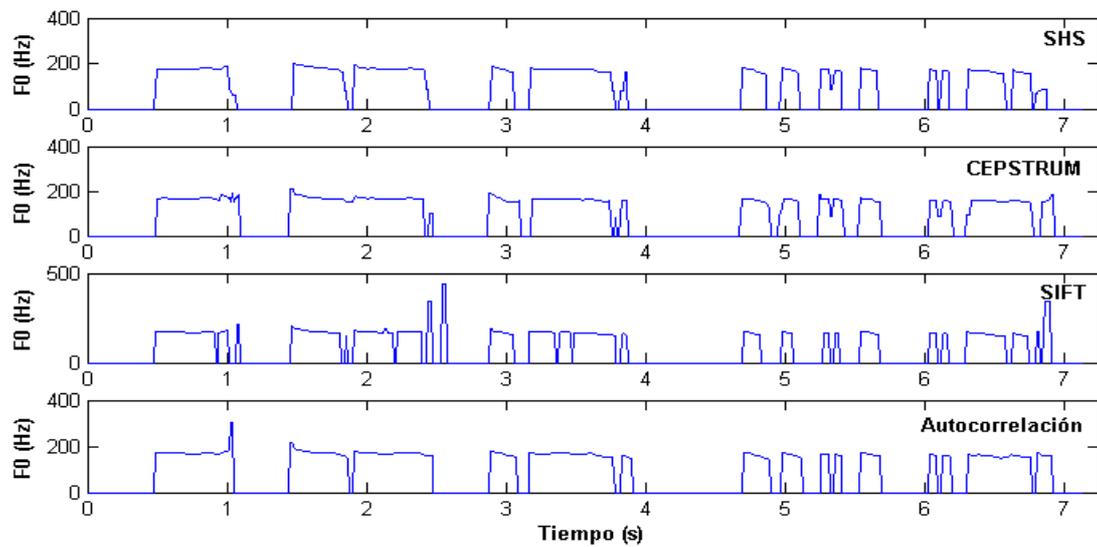


Fig.3.6. Contornos suavizados de “*Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever*”.

Algoritmos	EDT Sonoras	EDT Sordas
SHS	14 (1.95%)	0 (0%)
AUTOCORRELACIÓN	15 (2.09%)	0 (0%)
CEPSTRUM	17 (2.37%)	3 (0.42%)
SIFT	57 (7.95%)	14 (1.95%)

Tabla III. Errores en los contornos sin suavizamiento de “*Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever*”.

Algoritmos	EDT Sonoras	EDT Sordas
SHS	2 (0.28%)	0 (0%)
AUTOCORRELACIÓN	10 (1.9%)	0 (0%)
CEPSTRUM	7 (0.98%)	2 (0.28%)
SIFT	15 (2.09%)	10 (1.9%)

Tabla IV. Errores en los contornos suavizados de “*Well, he is nearly 93 years old and he still keeps his mind as swiftly as ever*”.

En la figura 3.7 se muestra el espectrograma de la oración *He dresses himself with an old black frock coat, usually several buttons missing*. La misma es pronunciada por un hombre que padece de disartria flácida y además dicho paciente presenta nasalidad que es un trastorno que afecta las medidas de entonación.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

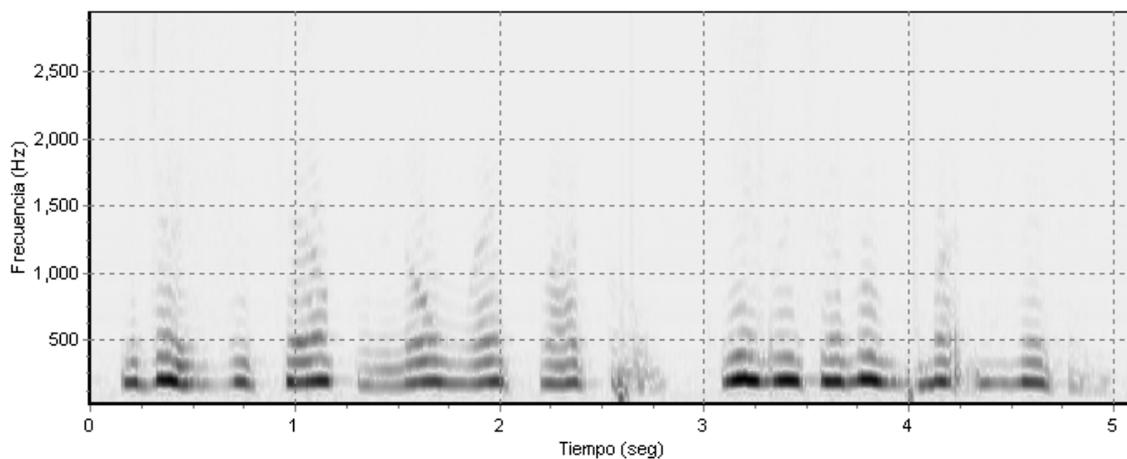


Fig.3.7. Espectrograma de la oración “*He dresses himself with an old black frock coat, usually several buttons missing*”.

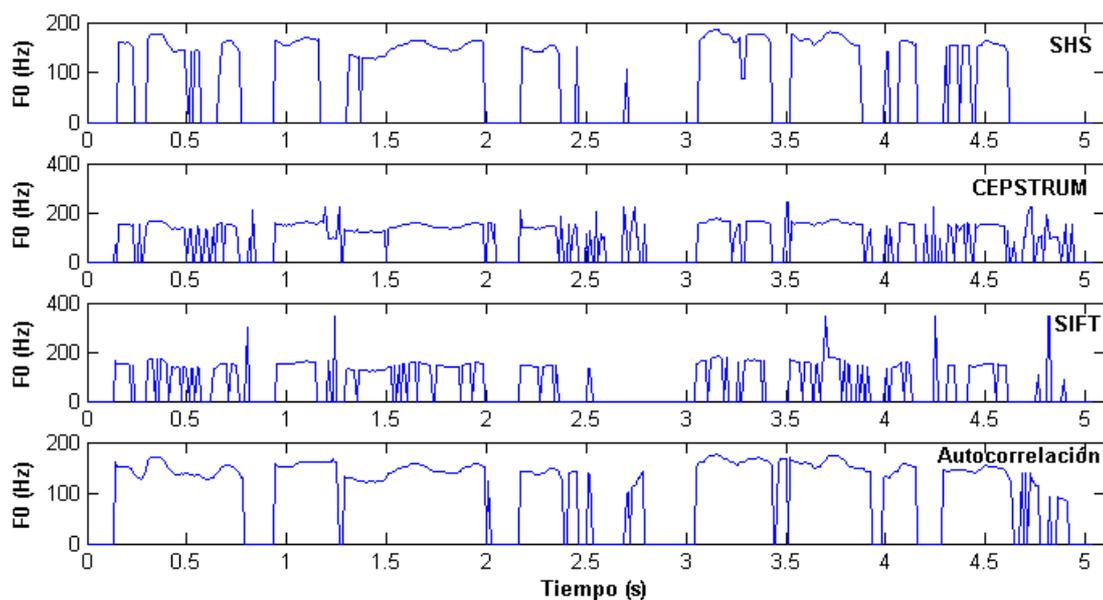


Fig.3.8. Contornos sin suavizamiento de la oración “*He dresses himself with an old black frock coat, usually several buttons missing*”.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

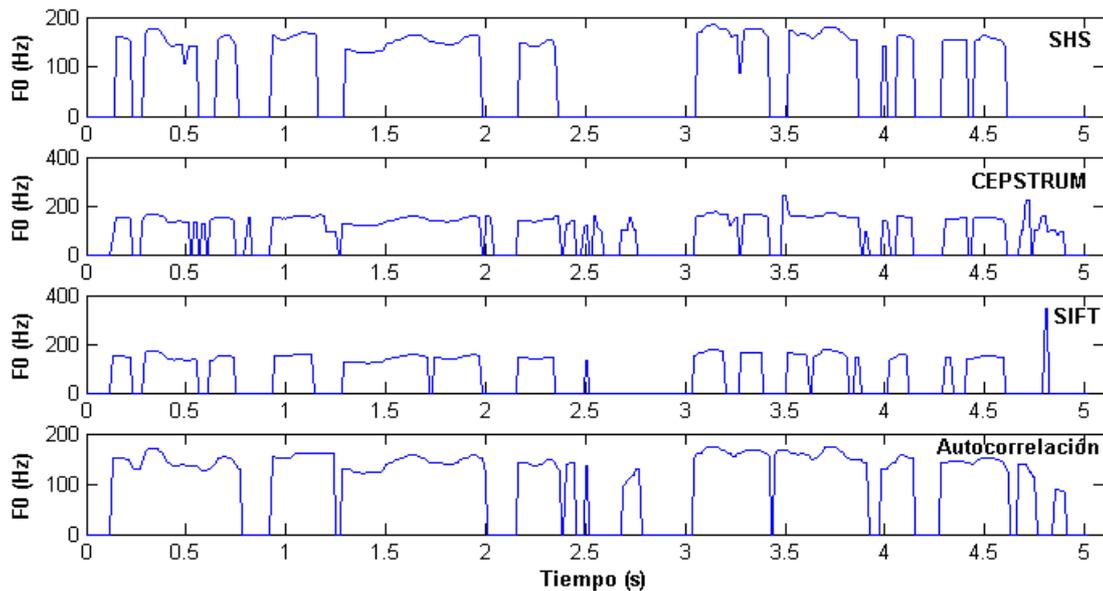


Fig.3.9. Contornos suavizados de la oración *“He dresses himself with an old black frock coat, usually several buttons missing”*.

Las figuras 7.8 y 7.9 presentan los contornos sin suavizar y suavizados de la detección de los algoritmos programados. Las tablas V y VI muestran los resultados de los análisis efectuados para determinar las fallas que ocurren en las mediciones.

Algoritmos	EDT Sonoras	EDT Sordas
SHS	18 (3.56%)	4 (0.79%)
AUTOCORRELACIÓN	8 (1.58%)	3 (0.59%)
CEPSTRUM	27 (5.35%)	9 (1.78%)
SIFT	35 (6.93%)	7 (1.39%)

Tabla V. Errores en los contornos sin suavizamiento de *“He dresses himself with an old black frock coat, usually several buttons missing”*.

Algoritmos	EDT Sonoras	EDT Sordas
SHS	10 (1.98%)	2 (0.4%)
AUTOCORRELACIÓN	4 (0.79 %)	2 (0.4%)
CEPSTRUM	18 (3.56%)	5 (0.99%)
SIFT	10 (1.98%)	2 (0.4%)

Tabla VI. Errores en los contornos suavizados de *“He dresses himself with an old black frock coat, usually several buttons missing”*.

Como se puede apreciar el algoritmo de autocorrelación tuvo menor cantidad de fallas en la detección de la entonación de este paciente. La nasalidad presente dificulta el cálculo puesto que la persona en el transcurso de la oración pierde mucho aire y la energía de la señal disminuye en gran medida. La figura 3.10 muestra una gráfica de la intensidad de la señal de voz, de ahí se puede observar que esta disminuye a gran escala con el tiempo, después de la pausa realizada alrededor de los tres segundos entonces se repite el mismo proceso hasta el final de la oración, logrando que algunos detectores fallen.

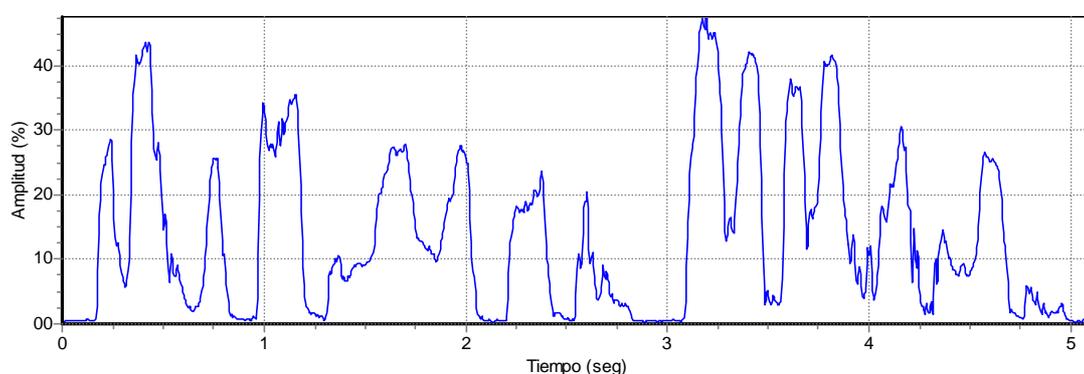


Fig.3.10. Gráfica de intensidad contra tiempo para el análisis de la oración *“He dresses himself with an old black frock coat, usually several buttons missing”*.

La principal causa de error consiste en el cambio de tramas sonoras por sordas. Con el SHS se observa un caso donde hay intervalos que no existen los armónicos suficientes para realizar una correcta detección de F0, sobre todo en los lugares donde la energía es muy pequeña.

3.2. Resultados generales.

Después de realizar este proceso a los 18 pacientes que se analizaron se obtuvieron los resultados que aparecen en las tablas VII y VIII, donde se procesaron más de 27000 tramas de señales del habla.

Se observa que el algoritmo de Hermes tuvo menor cantidad de fallas en la detección de los contornos de entonación de los pacientes. Las fallas del SHS ocurren principalmente sobre las medidas en los casos donde hay presencia de

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

Algoritmos	EDT Sonoras	EDT Sordas
SHS	572 (2.06%)	202 (0.73%)
AUTOCORRELACIÓN	673 (2.42%)	363 (1.31%)
CEPSTRUM	1580 (5.7%)	811 (2.93%)
SIFT	2526 (9.11%)	879 (3.17%)

Tabla VII. Errores en la detección de F0 sin suavizamiento.

Algoritmos	EDT Sonoras	EDT Sordas
SHS	251 (0.91%)	85 (0.31%)
AUTOCORRELACIÓN	496 (1.79%)	264 (0.95%)
CEPSTRUM	780 (2.83%)	487 (1.76%)
SIFT	1278 (4.63%)	605 (2.19%)

Tabla VIII. Errores en la detección de F0 con suavizamiento.

sonidos oclusivos sonoros, pues estos poseen poca cantidad de armónicos y no son suficientes para realizar una buena medición utilizando este método. En los pacientes con nasalidad existen intervalos con estas características.

Cuando las grabaciones se digitalizan, en la señal que se obtiene siempre existen un pequeño ruido aleatorio que pueden provocar que los detectores arrojen mediciones erróneas. En las condiciones de presencia de un ruido muy fuerte, los errores en de los detectores no se incluyeron en las tablas porque al haber presencia de señal fuerte demuestra la sensibilidad de los algoritmos. Estas causas provocaron errores gruesos principalmente, sobre todo en el algoritmo SIFT, en el cual la gran mayoría de las fallas fueron de este tipo. Es sorprendente que el algoritmo de Hermes no cometió errores gruesos con los pacientes en prueba lo que demuestra la robustez del algoritmo en presencia de ruido aleatorio.

En el anexo A aparecen tabulados los desempeños de los algoritmos por grupos disártricos. Se puede apreciar que hay pocas diferencias en cuanto al número de errores entre SHS y AUTOCORRELACIÓN, aunque con superioridad en alguno de los grupos. Luego del suavizamiento, es incuestionable la superioridad de SHS.

La desventaja fundamental del algoritmo de Hermes frente a los demás, es la complejidad en su implementación. Para la detección de la frecuencia fundamental se buscan agilizar el proceso de cálculo, pues esta resulta una medida básica en el procesamiento de voz. No obstante presenta gran precisión en el cálculo y es capaz de detectar cambios bruscos de la frecuencia producto a las enfermedades de los pacientes.

Conclusiones parciales

En el capítulo se han detallado los resultados y discutido las fortalezas y debilidades de los detectores analizados. La hipótesis planteada con relación al modelo auditivo se ha visto sustentada por los resultados expuestos.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

La incorporación del modelo auditivo en la detección de la frecuencia fundamental da lugar a un algoritmo robusto, que funciona mejor que otros detectores basados en el modelo de producción del habla, aplicado a voces patológicas.

El estudio comparativo realizado en los diferentes grupos disártricos muestra que el algoritmo de Hermes realiza la detección de periodicidad con la menor cantidad de fallas en la mayoría de las disartrias analizadas. No sucede así en las voces de los pacientes con disartria flácida donde el detector de autocorrelación muestra la detección más eficiente. El SIFT presenta el peor funcionamiento al contabilizar la mayor cantidad de errores en casi todas las detecciones realizadas.

De los algoritmos seleccionados e implementados el funcionamiento del SHS fue el más eficaz a la hora de conformar los contornos de entonación, los cuales son necesarios para la evaluación de la prosodia.

Recomendaciones

- Realizar investigaciones que conlleven a la implementación de algoritmos robustos para la detección de sonoridad que conjugados con el algoritmo de Hermes realicen una mejor detección de los contornos de entonación.
- Continuar la evaluación del algoritmo para determinar la cantidad óptima de subarmónicos para conformar de manera más precisa los contornos de entonación en voces patológicas.

REFERENCIAS BIBLIOGRÁFICAS

2. Aronson, A. E. (1993). *Dysarthria: Differential Diagnosis*. Mentor Seminars. Rochester. Michigan.
3. Baken, R. J. (1999). *Clinical Measurement of Speech and Voice*. San Diego: Singular.
4. Collins, M. (1984). "Integrating perceptual and instrumental procedures in dysarthria assessment". *Journal of Communications Disorders*, 5, 159-170.
5. Crystal, D. (1982). *Clinical linguistics*. Vienna: Sprienger-Verlag.
6. Darley, F. L., Aronson, A. E. & Brown, J. R. (1975). *Motor Speech Disorders*. Philadelphia: W.B. Saunders.
7. de Boer, E. (1977). Pitch theories unified. in *Psychophysics and Physiology of Hearing*, edited by E.F. Evans and J.P. Wilson (Academic, London), pp. 323-334.
8. Dubnowski, J. J., Schafer, R. W., y Rabiner, L. R. (1976). Real time digital pitch detector. *IEEE Trans. Acoust., Speech, Signal Process.* ASSP - 24: 2 - 8.
9. Duifhuis, H., Willems, L. F. y Sluyter, R. J. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustical Society of America*, 71: 1568 - 1580.
10. Ferrer, C. (2005). Cuantificación de parámetros de la voz para aplicaciones médicas. Tesis de Doctorado, Universidad Central de Las Villas, Santa Clara.
11. Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, 54: 1496 - 1516.

REFERENCIAS BIBLIOGRÁFICAS

12. Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America*, 83(1), 257-264.
13. Hernández-Díaz, M. E. (1996). Algoritmos para la extracción del período fundamental de la voz: Desarrollo y Evaluación. Tesis de Doctorado, Universidad Central de las Villas.
14. Hess, W. J. (1983). *Pitch Determination of Speech Signals - Algorithms and Devices*. Springer - Verlag, Berlín.
15. Hess, W. J. (1992). Pitch and Voicing Determination. in *Advances in Speech Signal Processing*, edited by S. Furui and M. Mohan Sondhi (Marcel Dekker, Inc., New York), pp. 3-42.
16. Houtgast, T. (1976). Subharmonic pitches of a pure tone at low S/N ratio. *Journal of the Acoustical Society of America*, 60, 405-409.
17. Houtsma, A.J. & Goldstein, J.L. (1972). The central origin of the pitch of complex tones: evidence from musical interval recognition. *Journal of the Acoustical Society of America*, 51, 520-529.
18. Kamath, N. (2005). Implementation, analysis and evaluation of some pitch detection algorithms. Indian Institute of Technology Bombay.
19. Kent R. D. and Ball M. J. (2000). *Voice Quality Measurement*. San Diego: Singular.
20. Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, 7, 7-23.
21. Kent, R.D., & Weismer, G. (1999). "Acoustic studies of disarthric speech: methods, progress, and potential".
22. Lehiste, I. (1970). "Suprasegmentals". Cambridge, MA: MIT Press.
23. Markel, J. D. (1972). The SIFT algorithm for fundamental frequency estimation. *IEEE Trans. AU - 20*: 367 – 377.
24. Martin, Ph. (1982). Comparison of pitch detection by cepstrum and spectral comb analysis. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP-82*, 180-183.

REFERENCIAS BIBLIOGRÁFICAS

25. Noll, A. M. (1967). Cepstrum pitch determination. *Journal of the Acoustical Society of America*, 41: 293 – 309.
26. Odhe, R. N., & Sharf, D. J. (1992). *Phonetic Analysis of Normal and Abnormal Speech*. Canada: Macmillan Publishing Company.
27. Rabiner, L. R., Cheng, M. J., Rosenberg, A. E. & McGonegal, A. (1976). A comparative study of several pitch detection algorithms. *IEEE Trans. Acoust. Speech, Signal Process.* ASSP - 24: 399 - 413.
28. Shriberg, L. D., & Kent, R. D. (1982). *Clinical Phonetics*. New York: Jhon Wiley & Sons.
29. Shrivastav, R., Sapienza, CH. M. (2003). Objectives measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America*, 114, 2217-2224.
30. Terhardt, E. (1979). Calculating virtual pitch. *Hear. Res.* 1, 155-182.
31. Terhardt, E.; Stoll, G. & Seewman, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71, 679-688.
32. Tiffany, W. R., & Carrell, J. (1977). *Phonetics: Theory and application*. New York: McGraw-Hill.
33. Zyski, B.J. & Weisiger, B.E. (1987). Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders*, 20, 367-378.

ANEXOS

ANEXOS

ANEXO A Tabla de resultados de los errores por grupos disártricos.

	SHS		CEPSTRUM		SIFT		AUTOCORRELACIÓN	
	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.
At	78 (1.77%)	22 (0.5%)	266 (6.03%)	68 (1.54%)	525 (11.89%)	137 (3.1%)	180 (4.08%)	33 (0.75%)
FD	90 (1.67%)	48 (0.89%)	227 (4.21%)	152 (2.82%)	405 (7.51%)	124 (2.3%)	36 (0.67%)	77 (1.43%)
KR	69 (1.18%)	42 (0.72%)	228 (3.89%)	222 (3.79%)	422 (7.2%)	212 (3.61%)	110 (1.88%)	105 (1.79%)
PK	153 (3.46%)	26 (0.59%)	419 (9.48%)	155 (3.51%)	500 (11.3%)	187 (4.23%)	147 (3.33%)	63 (1.43%)
SD	128 (2.65%)	44 (0.91%)	291 (6.01%)	116 (2.4%)	433 (8.95%)	101 (2.09%)	134 (2.77%)	58 (1.2%)
VT	54 (1.89%)	20 (0.72%)	149 (5.02%)	98 (3.54%)	241 (8.42%)	118 (4.26%)	66 (2.31%)	27 (0.97%)

Tabla IX. Errores de los algoritmos sin suavizamiento por grupos disártricos.

	SHS		CEPSTRUM		SIFT		AUTOCORRELACIÓN	
	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.	EDT Son.	EDT Sor.
At	14 (0.31%)	8 (0.18%)	96 (2.28%)	47 (1.07%)	250 (5.69%)	94 (2.14%)	131 (2.98%)	19 (0.43%)
FD	44 (0.81%)	20 (0.37%)	97 (1.8%)	61 (1.13%)	187(3.48%)	63 (1.17%)	20 (0.37%)	63 (1.17%)
KR	13 (0.22%)	15 (0.26%)	80 (1.37%)	140 (2.39%)	196 (3.35%)	147 (2.51%)	68 (1.16%)	72 (1.23%)
PK	94 (2.14%)	12 (0.27%)	247 (5.61%)	95 (2.16%)	341 (7.75%)	115 (2.61%)	111 (2.51%)	55 (1.24%)
SD	65 (1.35%)	21 (0.44%)	172(3.57%)	76 (1.58%)	190 (3.94%)	92 (1.91%)	107 (2.21%)	41 (0.85%)
VT	21(0.76%)	9 (0.33%)	88(3.19%)	68 (2.46%)	114(4.13%)	94 (3.4%)	59 (2.13%)	14 (0.51%)

Tabla X. Errores de los algoritmos suavizados por grupos disártricos.

Leyenda:

At → Disartria atáxica.

PK → Parkinson

FD → Disartria flácida.

SD → Disartria espástica.

KR → Korea.

VT → Temblor esencial.

EDT Son. → Errores en detección de tramas sonoras.

EDT. Sor. → Errores en detección de tramas sordas.

ANEXO B Listado del programa implementado en MATLAB ver 7.0.

```

function []=dcr_hermes1 ()

[filename, pathname] = uigetfile( '*.wav');
m = [pathname, filename];
[y,fs,NBITS,OPTS] = WAVREAD(m);

%-----
L_enventanado = 444; %--> Longitud de cada fragmento de la señal
f = enframe (y, L_enventanado,111); % Enventanar la señal (dividirla
% en ventanas), cada fila
% contiene los
% valores de las ventanas.

% Hasta aquí está estructurado el algoritmo en general.
% El siguiente paso es comprobar con una muestra para después
% implementarlo para todas las ventanas que se obtienen con la función
% enframe.

%-----
%% Tomo una fila de muestra para comprobar el algoritmo.

Muestra_prueba = f(50,:);
plot (Muestra_prueba)
Muestra_promedio=[]; % declarar una matriz para guardar el diezmo de
% la ventana de prueba.

%-----
%% Ciclo para promediar 4 muestras consecutivas.

for i=1:4:length(Muestra_prueba)-3
Muestra_promedio = [Muestra_promedio, mean(Muestra_prueba(i:i+3))];
end

%-----

Vent_hamm=hamming(111); % genero una ventana de Hamming de 111
% puntos que quedaron de los 444
x= Muestra_promedio .* Vent_hamm'; % Multiplico la señal con la
% ventana, el vector
% de la ventana tiene que ser
% conjugado porque
% al generarlo da un vector
% columna.
X = fft(x,256); % --> Calcular la FFT a 256 puntos.
XX = X(1:128); % --> Elimino la imagen.
mo_fft = abs(XX); %--> Calculo los valores absolutos de la FFT.
figure(2);

```

ANEXOS

```

plot(mo_fft) % Los ploteo.
%-----

%Este es el paso donde se enfatizan los picos.

Vabs = max(mo_fft); %cálculo del valor máximo de la FFT
Udb = (10 * log10(Vabs)) - 30; %cálculo del nivel umbral en db
Uabs = 10^(Udb/10); %busca el valor absoluto del umbral
[k, v] = findpeaks (mo_fft); %halla los picos y su localización
mo_fft1 = zeros (1,length(mo_fft));

for i=1:1: size (v,1) %Comparación con el nivel
    if v(i)> Uabs
        for j=-2:1:2
            mo_fft1(k(i)+j) = mo_fft(k(i)+j);
        end
    end
end

figure (3);
plot (mo_fft1);

%-----

% Realizo el filtrado de Hanning.

longitud = length (mo_fft);

Bn=zeros(length(mo_fft1)); % declaro la matriz Bn del mismo tamaño
% que mo_fft1
% para que no hallan confusiones
% respecto al tamaño.

Bn(1)=(0.5*mo_fft1(1))+(0.25*mo_fft1(2));
Bn(longitud)=(0.25*mo_fft1(longitud-1))+(0.5*mo_fft1(longitud));
for i=2:1:longitud-1
    Bn(i)=(0.25*mo_fft1(i-1))+(0.5*mo_fft1(i))+(0.25*mo_fft1(i+1));
end

%-----

% Cambio de escala.

n_muestras = 1:1:length(mo_fft1);
f_axis = n_muestras .* ((fs/4)/256);
figure(4);
plot (f_axis,Bn) % Ploteo en el eje de frecuencias.
%-----
% Esto es solo un ploteo en el eje de la frecuencia logarítmica.

s_axis = log2 (f_axis);
figure (5);

```

ANEXOS

```

plot (s_axis,Bn)                                %Ploteo en el eje s.

figure (6);
subplot (2,1,1), plot (f_axis,Bn);
subplot (2,1,2), plot (s_axis,Bn);

%-----
% Realizar la interpolación.

% 1ero es delimitar las octavas.
s_max = max (s_axis);
s_octava = [];

while s_max >= min(s_axis)
    s_octava = [s_max , s_octava];
    s_max = s_max-1;
end

% 2do es buscar los puntos donde tengo que interpolar.
s_interpolacion = [];
for i=1:1:length(s_octava)-1
    puntos = linspace(s_octava(i), s_octava(i+1), 48); % puntos por
                                                       % octava.
    s_interpolacion = [s_interpolacion, puntos(2:1:length(puntos)-1)];
end
% 3ro busco los puntos interpolados.
valores_interpolados = interp1(s_axis,Bn,s_interpolacion,'spline');

%Organizo el eje completo de la frecuencia con sus valores.
s_completa = [s_axis,s_interpolacion];
s_completa = sort(s_completa);
Af=[];
for i=1:1:length(s_completa)
    for j=1:1:length(s_axis)
        if s_completa(i)==s_axis(j)
            Af = [Af, Bn(j)];
        end
    end
    for k=1:1:length(s_interpolacion)
        if s_completa(i)==s_interpolacion(k)
            Af = [Af, valores_interpolados(k)];
        end
    end
end
end

%-----
%Multiplico por una función arc-tangente.

arc_tangent_window = atan(1:1:length(Af));
Pf = Af .* arc_tangent_window;
figure(7);
subplot (2,1,1); plot(s_completa, Pf)

```

ANEXOS

```

subplot (2,1,2); plot(s_completa, Af)

%-----
%Sumatoria de subarmonicos
for n=1:1:15
    hn=0.9^(n-1);
    desplazamiento=log2(n);
    s1=s_completa-desplazamiento;
    s=max(s1);
    s_auxiliar=[];
    Pf_final=[];
    for i=1:1:length (s_completa)
        if s_completa(i) <= s
            s_auxiliar=[s_auxiliar,s_completa(i)];
        end
    end
    Pf_final = interp1(s1, Pf, s_auxiliar, 'spline');
    Pf_fin(n,:) = [Pf_final, zeros(1,length(Pf)-length(Pf_final))];
    Pf_fin(n,:) = hn.* Pf_fin(n,:);
end

figure(8); hold on
plot(s_completa, Pf_fin(1,:));
plot(s_completa, Pf_fin(2,:), 'r');
plot(s_completa, Pf_fin(3,:), 'g');
plot(s_completa, Pf_fin(4,:), 'y');
hold off

figure(9);
subplot(6,1,1); plot(s_completa, Pf_fin(1,:));
subplot(6,1,2); plot(s_completa, Pf_fin(2,:));
subplot(6,1,3); plot(s_completa, Pf_fin(3,:));
subplot(6,1,4); plot(s_completa, Pf_fin(4,:));
subplot(6,1,5); plot(s_completa, Pf_fin(5,:));
subplot(6,1,6); plot(s_completa,
Pf_fin(1,:)+Pf_fin(2,:)+Pf_fin(3,:)+Pf_fin(4,:)+Pf_fin(5,:));

Hs=sum(Pf_fin);
figure(10);
plot(s_completa, Hs)
axis([min(s_completa) max(s_completa) (min(Hs)-0.01) (max(Hs)+0.1)]);
[V,S] = max(Hs);
s_pitch = s_completa(S);
f_pitch = 2^s_pitch;

```