Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Automática y Sistemas Computacionales

TRABAJO DE DIPLOMA

MODELACIÓN MATEMÁTICA DEL TRACTO VOCAL EN PRESENCIA DE *JITTER*.

Autor: Yosvel Tabares Alfonso.

Tutor: MSc. Roberto Díaz Amador.

Santa Clara

2013

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Automática y Sistemas Computacionales

TRABAJO DE DIPLOMA

MODELACIÓN MATEMÁTICA DEL TRACTO VOCAL EN PRESENCIA DE *JITTER*.

Autor: Yosvel Tabares Alfonso.

Tutor: MSc. Roberto Díaz Amador.

roberto@uclv.edu.cu

Santa Clara

2013

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 Tutor

Firma del Jefe de Departamento donde se defiende el trabajo

_____

Firma del Responsable de Información Científico-Técnica

PENSAMIENTO

Qué importa el tiempo sucesivo si en él hubo una plenitud, un éxtasis, una obra.

Jorge Luis Borges

DEDICATORIA

A mis padres por su cariño infinito, por estar conmigo siempre en las buenas y malas, por ser mi sostén y porque a ellos debo lo que soy…todo.

AGRADECIMIENTOS

Ante todo a mi familia, porque han resistido con fuerza mis veinticuatro años, dándome su amor infinito,

A Roberto por haber sido mi tutor, por apoyarme, sin él no me estuviese graduando hoy,

A mi hermano, porque es mi faro, por cumplir mis sueños y por tantas y tantas cosas…

A mis amigos de siempre: aquellos que han compartido conmigo fiestas, fines de semana de hambre, tristezas, decepciones; sobre todo, porque han sabido cómo ser amigos: Pototo, Roger, Asiel, Yosbel, Peña, Andro, Dayren,

A los que alimentaron mi espíritu y siempre me dieron la seguridad necesaria para graduarme…Adriana, Yisliany y Tuté,

A los profesores de la carrera por mi formación profesional.

TAREA TÉCNICA

1- Estudio de literatura científica actualizada sobre las afectaciones de la voz que se pueden modelar como *jitter*.

2-Estudio de los modelos de *jitter* existentes.

3- Creación de un modelo matemático para obtener la fuente de excitación en presencia de *jitter*.

4-Realización de experimentos que sirvan para evaluar el modelo propuesto.

5- Evaluación de los resultados utilizando señales simuladas y reales.

Firma del Autor

Firma del Tutor

RESUMEN

La medición de la calidad vocal presenta gran importancia en el diagnóstico y seguimiento de pacientes con diferentes patologías, asociadas no solo al aparato fonador, sino también de mayor complejidad afiliada al aparato neurológico. Dentro de las medidas de calidad más utilizadas es la medición de la periodicidad de la voz, la que tiene un reto muy investigado, pero aun abierto, en la presencia de anomalías en la periodicidad como pueden ser el jitter, el shimmer o la presencia de ruidos. Uno de los factores interesantes, aun no completamente resuelto, es evaluar el desempeño de un modelo que permita recuperar la información de la fuente de información separándola del filtro en presencia de distintos niveles de afectación de la voz. El trabajo pretende evaluar el modelo de predicción lineal para recuperar la fuente de información y con ella los instantes de cierre glotal en presencia de jitter por duración y por posición. Adicionalmente se evalúa el comportamiento del modelo propuesto utilizando una base de datos de señales reales. Los resultados de este trabajo apuntan que, el modelo presenta mejor desempeño en presencia de jitter por posición que en presencia de jitter por duración, para lo cual se utilizó como medida de calidad la diferencia entre los instantes glotales de la señal glotal recuperada por el modelo y la señal que se utilizó para generar una vocal sintética. En el trabajo con señales reales se utilizó como medida de calidad una comparación de la señal recuperada y la información que presenta la base de datos utilizada.

TABLA DE CONTENIDOS

PENSAMIENTO i

DEDICATORIA ii

AGRADECIMIENTOS iii

TAREA TÉCNICA iv

RESUMEN v

TABLA DE CONTENIDOS vi

INTRODUCCIÓN 1

CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA 4

1.1 Calidad Vocal. 4

1.2 Aplicaciones de la Calidad Vocal. 5

1.3 Jitter y Shimmer. 5

1.4 Recuperación de la fuente de excitación Glotal. 7

1.5 Diferentes Modelos de Jitter. 9

1.6 Detención de F0 en presencia de jitter. 10

1.7 Conclusiones del Capítulo 11

CAPÍTULO 2. MATERIALES Y MÉTODOS 12

2.1 Señales sintéticas 12

2.2 Síntesis de vocal sostenida perturbada por jitter (modelos E/C y S/A) 16

2.3 Base de datos de señales glotales reales 16

2.4 Creación de un modelo del tracto vocal utilizando la Predicción Lineal 19

2.5 Descripción de los experimentos realizados. 20

2.6 Conclusión del Capítulo 22

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN 23

3.1 Evaluación utilizando señales sintéticas. 23

3.1.1 Evaluación cuando no ocurre Jitter. 23

3.1.2 Evaluación en presencia de Jitter por posición. 24

3.1.3 Evaluación en presencia de jitter por duración. 25

3.1.4 Comparación de la influencia de los dos modelos de jitter tratados utilizando 26

como línea base de la evaluación cuando no ocurre jitter. 26

3.2 Evaluación utilizando voz grabada. 27

3.2.1 Obtención de la función de transferencia del tracto vocal. 27

3.2.2 Evaluación en señales reales. 28

3.3 Conclusiones del Capítulo 31

CONCLUSIONES Y RECOMENDACIONES 32

Conclusiones 32

Recomendaciones 33

REFERENCIAS BIBLIGRÁFICAS 34

INTRODUCCIÓN

El habla es de gran importancia para el hombre, puesto que es reconocida como la principal vía de comunicación entre los seres humanos. Sin embargo, lo que muchos desconocen es que constituye un proceso sumamente complejo, donde intervienen varios órganos y sistemas. La afección de alguno de estos órganos o sistemas repercute en la forma en que se genera y percibe el habla.

El campo de las aplicaciones médicas del procesamiento de voz constituye, aún, un sitio poco explorado por lo que existe hacia la temática, gran interés por parte de la comunidad científica internacional. Puede decirse que, resultan variados los campos desde los cuales es abordada la información contenida en la señal de voz, pueden mencionarse: los sistemas de comunicación con transmisión, análisis, síntesis y reconocimiento de voz; en los estudios de fonética acústica y en la lingüística, centralizados en la entonación; de vital importancia es su aplicación en la enseñanza de sordos, entre otras. En este caso concreto, resulta de gran utilidad la obtención de medidas capaces de indicar la severidad de las manifestaciones patológicas del habla, lo que permitiría efectuar diagnósticos más acertados cuando de patologías del habla se trata, así como documentar de forma objetiva la evolución del paciente y la efectividad del tratamiento.

La influencia de afecciones como: disfonía, aspereza, la presencia de ruido de aire en la voz, voz soplada, entre otras**,** provocan que el habla porte no sólo el mensaje (estados emocionales o de ánimo y procedencia social de un individuo) sino también información útil desde el punto de vista clínico. Las perturbaciones típicas incluyen alteraciones de la amplitud (*shimmer*) y frecuencia (*jitter*) de los pulsos. La perturbación que impone mayores dificultades a las suposiciones de periodicidad es el *jitter*, por lo que el estudio de su influencia y los modos de reducirla resultan de particular interés.

Las medidas de Calidad Vocal determinan perturbaciones en fonemas (sonidos individuales). A medida que se incrementa la complejidad de las unidades del lenguaje a analizar (desde los fonemas a las oraciones transitando de calidad vocal a prosodia) se incrementa también, la complejidad del desarrollo y la obtención de mediciones correspondientes. Este trabajo se mantiene en la línea de desarrollo de mediciones de Calidad Vocal en segmentos sonoros donde la señal es aproximadamente periódica. Los modelos tienen implicaciones diferentes en el modo en que se afecta la señal y en el rendimiento de los algoritmos diseñados para medir las perturbaciones de la periodicidad (*jitter*).

Como antecedentes directos del presente estudio se encuentran las investigaciones: *Conversión de Texto a Voz para Implante Coclear* de Roberto Díaz Amador; *Influencia de Modelos de Jitter en Medidas de Perturbación de la Periodicidad* de Alexi José González Hernández. Esta última constituye el antecedente más cercano, en la misma se evaluó el comportamiento de métodos reportados de estimación de duración y amplitud de los pulsos glotales, ante la presencia de niveles patológicos de *jitter*, considerando dos modelos de perturbación del pulso. Aun cuando parezca que existen puntos en común entre la investigación antes citada y la que se expone, puede decirse que presentan grandes puntos divergentes en cuanto a los modelos de realización.

Lo planteado anteriormente ha permitido trazar el siguiente tema de investigación: *Modelación matemática del tracto vocal en presencia de jitter.*

La investigación en cuestión posee gran **novedad** e importancia. Este estudio propiciará conocimientos en el área de procesamiento de señales. Además representará, desde el presente inmediato a la investigación, un beneficio social y económico para el país. Los resultados obtenidos son de aplicación para el desarrollo de investigaciones en el campo de la calidad vocal, la logopedia y la foniatría. También los resultados obtenidos servirán de material de estudio para estudiantes y profesores de la facultad de Ingeniería Eléctrica.

Para complementar el tema propuesto se planteó el siguiente **problema científico**:

¿Cómo contribuir a la creación de un modelo matemático para obtener la fuente de excitación en presencia de jitter?

Para dar solución al problema antes mencionado se trazaron los siguientes objetivos:

Objetivo General: Modelar el trato vocal en presencia de jitter utilizando el modelo de predicción lineal para obtener la fuente de excitación glotal mediante filtrado inverso.

Objetivos específicos:

1. Obtener el modelo matemático de predicción lineal para simular el comportamiento del trato vocal en presencia de jitter.
2. Evaluar el comportamiento del modelo a diferentes niveles de afectaciones de la periodicidad bajo condiciones controladas.
3. Evaluar el modelo obtenido utilizando bases de datos que contengan la señal electroglotográfica para validar su capacidad de dar seguimiento a los puntos de interés de la señal electroglotográfica (EGG).

El informe quedó estructurado en tres capítulos. El primero contiene la conceptualización de términos imprescindibles en la realización del estudio tales como: calidad vocal, *Jitter* y *Shimmer*, así como diferentes modelos de *Jitter*: por posición y por duración. Además se brinda una panorámica general en torno al problema abordado.

En el capítulo 2: *Materiales y métodos* se hace una breve descripción de los métodos que se utilizan para la creación de un modelo del tracto vocal a partir de señales sintéticas y señales reales. En el caso de las señales sintéticas se evalúa el desempeño del modelo en presencia de diferentes niveles de jitter. En el caso de las señales reales se realiza una comparación entre la salida del modelo y la forma de onda glotal.

Los resultados obtenidos en los experimentos y la realización del análisis se encuentran expuestos en el tercer capítulo, planteando las implicaciones de los mismos en cada caso. Finalmente, se recogen en el informe las conclusiones, las recomendaciones y las referencias bibliográficas.

CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA

El habla, como medio de comunicación humana, expresa gran información en varios niveles: lingüístico, para-lingüístico y extra-lingüístico [1]. En el primer nivel contiene la información de lo que se desea emitir (el "texto" del mensaje). El segundo porta información sobre los estados de ánimo, emocionales u otros, de la persona, así como su procedencia social y otros factores culturalmente determinados. El tercer nivel alcanza todos los factores físicos y fisiológicos (incluyendo todos los aspectos orgánicos involucrados en la producción del habla) característicos del locutor.

Dada la importancia que porta el habla para los seres humanos, debido a que es la vía fundamental de comunicación, se comprende la amplitud de las investigaciones que se dedican al estudio del habla en estos tres niveles. En este trabajo se abordan técnicas para analizar la información extra-lingüística del habla.

1.1 Calidad Vocal.

Para valorar perceptiblemente la calidad vocal, se dispone de un referente innegable en la escala GRBAS [2]. El sistema GRBAS está formado por cinco parámetros: G (grado de disfonía), R (Roughnes: Rugosidad de la voz), B (Breathiness: respiración dificultosa), A (Aesthenia / astenia o grado de fatiga de la voz), S (Strain o grado de tensión vocal). El sistema de puntuación de cada parámetro va desde el 0 al 3. Siendo 0 normalidad, 1 ligera alteración, 2 alteración moderada y 3 alteración severa. Una variedad de escalas han sido creadas para intentar evaluar la calidad de la voz pero en comparación con otras, la escala GRBAS tiene la ventaja de poder ser usada diariamente gracias a su simplicidad [3].

1.2 Aplicaciones de la Calidad Vocal.

En la calidad vocal se realizan aplicaciones sobre aspereza de la voz que han sido relacionadas con las perturbaciones de la periodicidad de los pulsos glotales y pueden dividirse en perturbaciones de frecuencia (jitter) y de amplitud (shimmer). Estas perturbaciones de amplitud y frecuencia son provocadas por la vibración irregular de las cuerdas vocales, y rara vez aparecen de forma independiente, también está el jadeo que se evalúa el nivel de severidad en una escala de cuatro puntos donde al primero se le otorga el valor 0 que responde a la ausencia de disfonía y los valores 1, 2 y 3 indican los niveles ligero, moderado y severo en cada uno de los parámetros valorados. Se trata de una escala perceptiva que es conocida y empleada de forma habitual en todo el mundo a pesar de que en Cuba no todos los profesionales ORL y logopedas la usen aún. El presente estudio considera que se trata de la prueba estándar para determinar la validez de otros instrumentos de medida de la calidad vocal**.** Se toma como referencia el protocolo básico de la Sociedad Europea de Laringólogos [4] donde se considera que las medidas de perturbación de la amplitud y de la frecuencia (shimmer y jitter) y las de relación entre la señal y el ruido son las más robustas para cuantificar las características perceptuales de la calidad vocal.

1.3 Jitter y Shimmer.

Jitter: este parámetro estima la variación del período de vibración de la glotis entre dos períodos consecutivos, en una modulación de la periodicidad del signo de la voz [5]. El Jitter es otro parámetro de la voz muy utilizado, que mide el grado en que los ciclos son distintos entre sí en lo que respecta a su duración, o periodo (Figura 1.3) [6][7][8]. Si los ciclos fueran idénticos unos a otros, el jitter sería cero. Esto no ocurre nunca en la voz humana, donde siempre hay pequeñas variaciones de un ciclo a otro. Sin embargo, las variaciones son tan pequeñas que el jitter se mide en microsegundos. Cuando los ingenieros sintetizan voz artificial por computador, la voz suena robótica porque, entre otras razones, los ciclos son idénticos unos a otros. Para evitar esto, introducen en sus fórmulas un factor de error aleatorio que crea pequeñas diferencias entre los ciclos y, de este modo, la voz suena más natural. En el otro extremo, las voces patológicas por diversas etiologías (enfermedades neurológicas, pólipos, nódulos, tumores, parálisis de una cuerda vocal),

suelen tener jitter altos porque los ciclos son muy distintos entre sí a consecuencia de las irregularidades de vibración de las cuerdas vocales. Las voces con jitter altos suenan "ásperas" y desagradables al oído humano. El tabaquismo, además de bajar la frecuencia fundamental (F0) también afecta el jitter, incrementándolo por encima de sus valores normales.

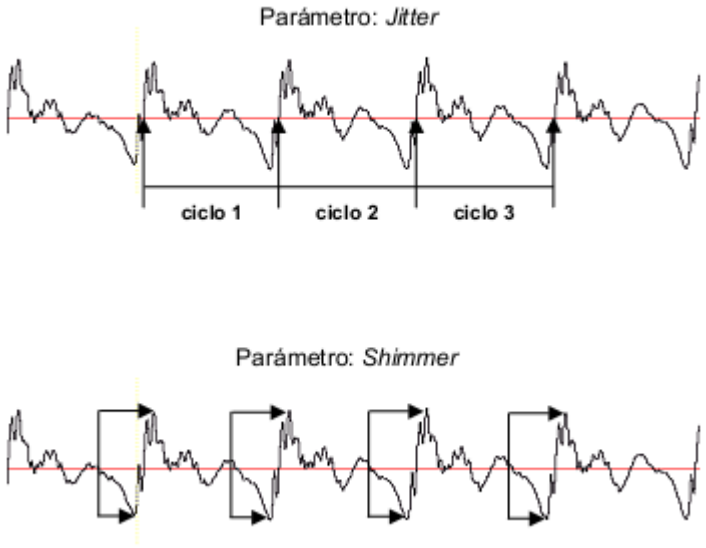


Figura 1.3. Parámetros útiles para diagnosticar voces patológicas: Jitter y Shimmer. El Jitter da una medida de cuán distintos son los ciclos respecto a su duración (periodos). El Shimmer mide cuán distintos son los ciclos respecto a su amplitud máxima, pico a pico.

Técnicamente, el *jitter* es el promedio de las diferencias entre los periodos de medidas ciclo a ciclo. Se calcula de forma semiautomática mediante un software específico a partir de una muestra voz que consiste en la **fonación sostenida de una vocal** (usualmente /a/) durante 1.5 - 3 segundos. El algoritmo de cálculo debe seguir los siguientes pasos:

1. Identificar los ciclos de la voz (es la parte más compleja, porque es fácil que el software cometa errores).

2. Calcular los periodos, o duración de cada ciclo.

3. Restar cada periodo del anterior y hallar el promedio de las diferencias. El resultado se da en microsegundos. Un valor en torno a 80 microsegundos se considera un límite superior de normalidad [9]. Las voces patológicas tienden a superarlo. En términos relativos, supone aproximadamente un 1% del periodo de un ciclo.

Existen diferentes variantes para calcular el Jitter que consisten en calcular las diferencias, no ciclo a ciclo, sino entre grupos de ciclos consecutivos que se promedian previamente. Dependiendo del número de ciclos que se promedian antes de calcular las diferencias, se obtienen distintos parámetros que miden rasgos específicos de la voz [9].

Shimmer: Este parámetro estima la variación de la máxima amplitud de la fonación entre dos períodos consecutivos de vibración de la glotis [10]. Variación de la amplitud de la onda ciclo a ciclo, se basa en el grado de correlación que tienen los picos periódicos de la función de autocorrelación [8][7][11] .*Shimmer*, que sería equivalente al jitter, pero esta vez en relación a la amplitud del ciclo, en lugar de su duración. Se calcula a partir de la amplitud máxima de cada ciclo, medida pico a pico, o distancia entre el pico positivo más alto y el negativo más bajo (Figura 1.3). Da una idea del grado de disparidad que existe entre las amplitudes de los ciclos consecutivos. De nuevo, si todos los ciclos fueran iguales, el shimmer sería 0 y la voz sonaría poco natural. Las voces humanas siempre tienen pequeñas variaciones entre las amplitudes de sus ciclos que dan lugar a shimmers superiores a cero. El valor que se considera como límite superior de normalidad se sitúa sobre 0,35 decibelios, o en torno al 4 % de la amplitud total de un ciclo [9] . Las voces patológicas tienden a sobrepasar estos umbrales.

También el *shimmer* presenta varias versiones si las diferencias se calculan, en vez de ciclo a ciclo, promediando grupos de ciclos consecutivos. A modo de resumen hay que destacar que, dada la gran variabilidad que la voz humana presenta de forma natural, estos parámetros también están sometidos a grandes variaciones dentro de la normalidad. El diagnóstico de una voz patológica siempre es aproximativo y gana peso cuando son varios los parámetros que confluyen en valores anormales.

1.4 Recuperación de la fuente de excitación Glotal.

El proceso de obtención de la fuente de información glotal puede pasar en primer lugar, por obtener directamente esa información utilizando un electroglotógrafo, o recuperando la información de la fuente a partir de la señal de voz. Los instantes de cierre y apertura glotal contienen información muy importante sobre la señal de voz como la obtención de la frecuencia fundamental F0, la obtención de la forma de onda de glotal (g(t)) y otras, esto no es un problema sencillo. Algunas formas de medición directa en sujetos vivos han sido

altamente invasivas debido a la colocación de transductores. Otras basadas en el cálculo del área de las cuerdas vocales como la electroglotografía [12] y laringoscopia [13] han fallado debido a la conductividad eléctrica transglotal y los requisitos muy estrictos con la configuración instrumental para obtener una secuencia de imágenes adecuada para medir el área, respectivamente.

Ante las dificultades que se reportan en la estimación de la señal de excitación (ya sea experimental directa o mediante modelos físicos) ha sido frecuente la búsqueda de alternativas con menor basamento físico y mayor abstracción matemática. El método más usado es el empleo de la señal residual del filtrado inverso para estimar la forma de onda de g(t), aunque el filtrado inverso tiene aparejadas una serie de limitaciones que se tornan severas en voces patológicas. En este último caso el problema es separar adecuadamente la información de la fuente de la información del filtro. Las formas de onda obtenidas han sido representadas mediante modelos paramétricos donde se le asignan determinadas funciones matemáticas a las fases de apertura y cierre de la forma de onda de g(t) (ver Figura 1.4.). Entre estos modelos los más empleados son los polinomiales y trigonométricos de Rosemberg [14] y el modelo de Liljencrants-Fant [15].

En estos se presentan tres métodos de aplicar el filtrado inverso a voz sintética utilizando el modelo de Liljencrants-Fant para recuperar la información de la fuente aplicando el método de Prony y el método del gradiente descendente [16]. En ese trabajo se evalúan los resultados utilizando voz sintética (particularmente vocal I sostenida) y utilizando voz grabada, pero en ningún caso se tiene en cuenta la presencia de dificultades en la periodicidad de la señal como jitter o shimmer, solamente se valora la presencia de diferentes niveles de ruido. Sakakibara [17] concibe un modelo multiparamétrico del flujo glotal que incluye además la estructura supraglotal y laríngea, y los resultados son evaluados por comparación con un flujo simulado de señal periódica. En este caso tampoco se tiene en cuenta la presencia de desviaciones en la periodicidad de la señal y no se realizan comparaciones que impliquen voz grabada. Otro trabajo interesante es la detección de los instantes de cierre glotal que se presentan [18], basados en el hecho de que lo más importante no es la detección exacta del flujo glotal, sino la detección de los instantes de cierre, ya sean bruscos o no. En este caso utilizan representaciones tiempo-escala (transformada wavelet), sin embargo, aunque puede ser utilizado para seguimiento del pitch

en el trabajo no se hace una evaluación del método con esta finalidad, por lo que tampoco se tienen en cuenta su comportamiento frente a variaciones de la periodicidad. En ninguno de los casos que el autor ha podido investigar se ha tenido en cuenta la influencia de variaciones de la periodicidad en la obtención de la información de la fuente glotal o su derivada, por lo que en este trabajo se intenta evaluar la influencia del jitter para recuperar la fuente glotal utilizando como modelo de fuente el descrito por Rosemberg 1971 y como modelos de jitter por posición (modelo S/A) [19] y el modelo de jitter por duración (modelo E/C) [14].

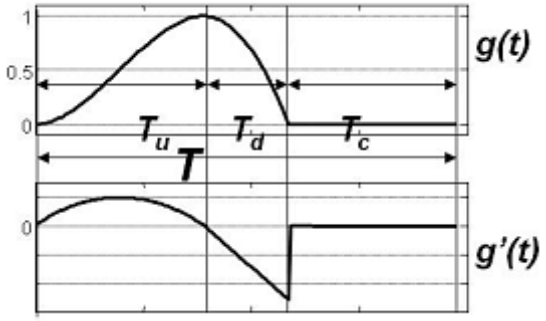
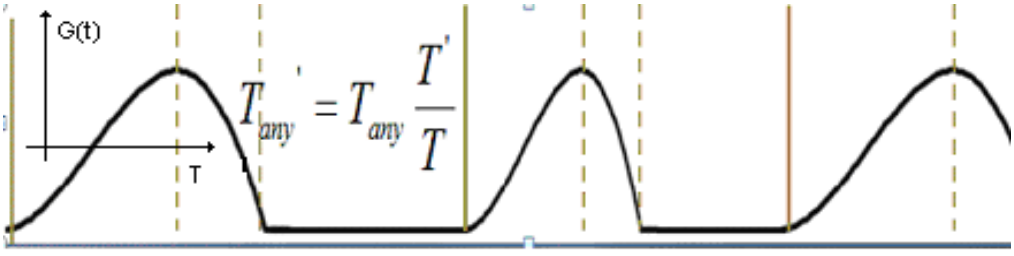


Figura 1.4. Formas de onda y fases de la señal de excitación glotal g(t) y su derivada g'(t). *Tu* es la duración de la fase de apertura, *Td* la del cierre, *Tc* la fase cerrada, y T la duración total del pulso.

1.5 Diferentes Modelos de Jitter.

Se han utilizado dos modelos de *Jitter*, por una parte el que concibe las perturbaciones de duración como contracciones o expansiones de los pulsos (Contracción/Expansión) y por otra el que las concibe como acercamientos y separaciones de los mismos (Acercamiento/Separación de los pulsos.) (Ver Figura 1.5.).



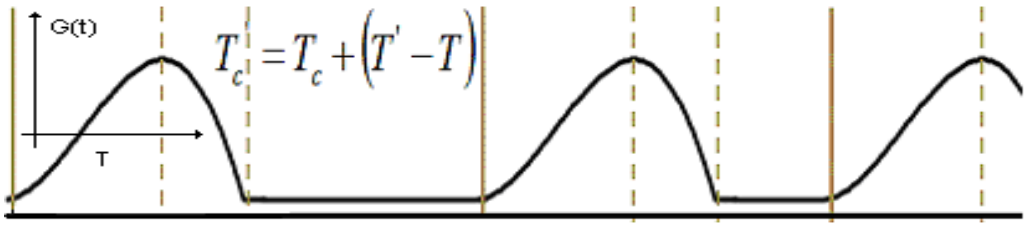


Figura1.5. Modelos de Jitter: Arriba, Contracción/Expansión de los pulsos, Abajo Acercamiento/Separación de los pulsos.

El modelo de jitter S/A contempla el paso de un tren de impulsos por un filtro paso bajo, esto provoca que aparezca un pulso en las posiciones de los impulsos de Dirac, la forma del pulso (fase abierta) siempre es constante [19]. Las variaciones de *jitter* que se modelan mediante el modelo de E/C, asume constantes las relaciones de duración entre las fases se asumen constantes y el inicio del pulso se asume en el instante de cierre [14].

Es de destacar, que no se han encontrado trabajos donde se evalúen simultáneamente ambos modelos. En los artículos disponibles ni siquiera se considera la posibilidad de la existencia de un modelo alternativo al que se asume en el trabajo en cuestión [6].

1.6 Detención de F0 en presencia de jitter.

El periodo del pitch o su inverso, frecuencia fundamental (F0), se determina por la velocidad de apertura o cierre de las cuerdas vocales en la laringe durante la fonación de sonidos sonoros, es el sinusoide de mayor contenido energético cuyo inverso corresponde, a su vez, al periodo fundamental (T0). En forma general, la definición de F0 se determina para intervalos infinitos de análisis; sin embargo, para efectos prácticos, su estimación se realiza sobre intervalos finitos, que permitan cubrir varios periodos del pitch (en este caso F0 se calcularía como el valor promedio de todos los F0 para un determinado intervalo). Otra forma de calcular el pitch es de manera instantánea, es decir, a partir de la diferencia entre dos GCI (*glotal close instant*) consecutivos [20][21], el jitter se corresponde a las variaciones de F0 que existen en el tramo de habla analizado, representadas como un ruido por modulación en frecuencia.

1.7 Conclusiones del Capítulo

Luego de haber realizado el primer capítulo, puede concluirse que para valorar la calidad vocal, se utiliza la mayoría de las veces el referente de la escala GRBAS; atendiendo a disfonía, aspereza, la presencia de ruido de aire en la voz, voz soplada, la debilidad o abstinencia tímbrica. Dentro de las aplicaciones de la calidad vocal se realizan estudios sobre aspereza de la voz, que ha sido relacionada con las perturbaciones de la periodicidad de los pulsos glotales y pueden dividirse en perturbaciones de frecuencia (jitter*)* y de amplitud (shimmer*)*. Existen dos modelos de *Jitter* que conciben las perturbaciones de duración como contracciones o expansiones de los pulsos. Dentro de los modelos de uso del jitter se encuentra el uso del modelo S/A (*jitter* por posición) y el E/C (*jitter* por duración).

La definición de F0 se determina para intervalos infinitos de análisis, que permitan cubrir varios periodos del pitch (en este caso F0 se calcularía como el valor promedio de todos los F0 para un determinado intervalo). Otra forma de calcular el pitch es de manera instantánea, es decir, a partir de la diferencia entre dos GCI (*glotal close instant*) consecutivos.

CAPÍTULO 2. MATERIALES Y MÉTODOS

En este capítulo se emplearán los modelos de jitter, tanto las perturbaciones de duración como las contracciones o expansiones de los pulsos (Contracción/Expansión) y acercamientos y separaciones de los mismos (Acercamiento/Separación de los pulsos), para la creación del tracto vocal a partir de las señales sintéticas y señales reales. En las señales sintéticas se evalúa el desempeño del modelo en presencia de diferentes niveles de jitter. En las señales reales se realiza una comparación entre la salida del modelo y la forma de onda glotal.

2.1 Señales sintéticas

Para sintetizar las señales empleadas se usó el MatLab 7.4, a partir de un algoritmo que permite sintetizar ambas señales. De acuerdo al modelo utilizado el programa genera la forma de onda glotal, la cual es la convolución y la función del tracto vocal, que es fijo para cada modelo y como resultado se obtienen las señales sintéticas, produciéndose una vocal sostenida "a". Esto se conoce como el modelo "Fuente – Filtro" de producción de la voz. En la Figura 2.1.1 se muestra un esquema del modelo Fuente – Filtro.

Si g(t) representa los pulsos de aire generados en la glotis, en la Figura 2.1.1 denotado como pulso glotal, que son conformados espectralmente por la función de transferencia del tracto vocal H(f) y por la radiación de los labios R(f), la señal de voz se puede expresar como:

s (t)=g (t)*h(t)*r(t) donde ' * ' denota la operación de convolución

O aplicando la transformada de Fourier:

S(f)= G(f) H(f) R(f)

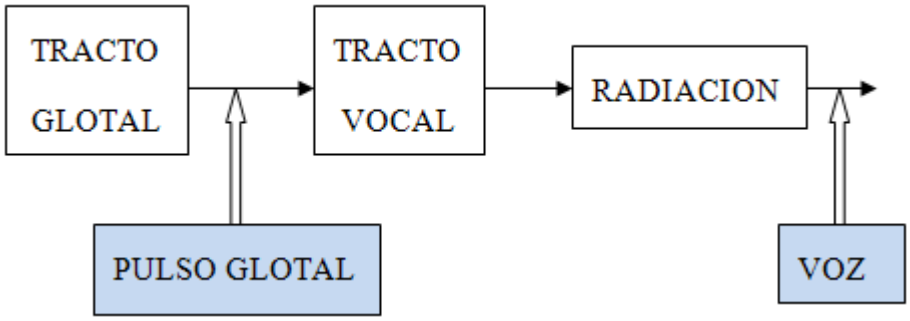


Figura 2.1.1. El modelo Fuente –Filtro.

En general, para simplificar este modelo se supone que el tracto glotal está formado por dos polos de los cuales uno se compensa con los efectos de radiación de los labios[15][22]

$(R\ (z) = 1-\alpha z\ -1)$ y el otro mediante un preénfasis que se hace a la señal de la voz antes de ser analizada, por lo que se funden las características de frecuencia de la fuente (características pasa-bajos) y el radiador (características pasa-altos) en una señal de espectro plano en la glotis (fuente), de manera que toda la conformación de la señal acústica pueda atribuirse al tracto vocal. Otra alternativa es considerar el radiador como una operación de derivación, que convierte una señal de velocidad de flujo g (t) en una señal de presión s (t). En la Figura 2.1.2 se muestra el modelo fuente-filtro simplificado.

Figura 2.1.2. El modelo Fuente - Filtro simplificado.

Las cuerdas vocales en vibración constituyen una fuente periódica (sonidos sonoros), ya que producen una onda sonora periódica, si, al contrario, la fuente se encuentra en la cavidad bucal, la señal producida será de tipo aperiódico (sonidos sordos) , pues tendrá como base una fricción producida por un estrechamiento del canal por el cual pasa el aire (sonidos fricativos) o bien una oclusión como consecuencia del cierre total de dicho canal (sonidos oclusivos), el resultado acústico de una fuente aperiódica se conoce como ruido.

Debido a que el habla es un fenómeno continuo, su función de transferencia varía a lo largo del tiempo. Partiendo de esta idea básica, es posible esquematizar la producción de los

sonidos del habla, teniendo en cuenta que existe también la posibilidad de combinar la fuente periódica (sonidos sonoros) y la aperiódica (sonidos sordos). En la Figura 2.1.3 se representa este esquema.

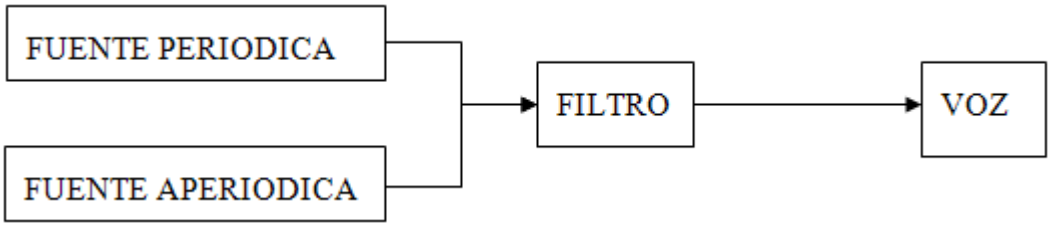


Figura 2.1.3. Esquema de producción del sonido.

Para generar la señal de excitación glotal, se puede utilizar una gran variedad de métodos, que se han orientado al modelado físico del sistema vibratorio de la glotis [22][23] . Sin embargo; no existe consenso en cuanto a la cantidad de elementos a considerar en los modelos, y los resultados son limitados en cuanto a la variedad de modos de vibración que permiten explicar [24][25] . Pero lo más frecuente ha sido la búsqueda de alternativas con mayor abstracción matemática aún a costo de un menor fundamento físico, como es el caso del empleo de la señal residual del filtrado inverso para estimar la forma de onda de excitación. Las formas de onda obtenidas se han representado mediante modelos paramétricos que le asignan determinadas funciones matemáticas a las fases de apertura y cierre de la glotis, entre estos modelos se destacan los poligonales y trigonométricos descritos por Rosemberg, y el modelo de Liljencrants-Fant. Los criterios de naturalidad de la voz sintetizada con estos modelos y la minimización del error medio cuadrático respecto a la señal residual han sido ampliamente descritos [5][14][15][27] y basado en esos resultados algunos autores [28] plantean que hasta el momento no existen razones para preferir uno u otro modelo de excitación. Por otra parte, una serie de autores manifiestan que hasta el momento no existen razones para preferir uno u otro modelo de excitación, por lo que en este trabajo el modelo para la síntesis de la forma de onda glotal que se utilizó fue el modelo de Rosemberg , que se muestra en la Figura 2.1.4 que se define por la ecuación (2.0).

$$exc = \frac{\alpha}{2}\left[1 - \cos\left(\frac{t}{tp}\pi\right)\right] \quad \text{con } 0 \le t \le tp \qquad (2.0)$$

$$exc = \frac{\alpha}{2}\left[1 + \cos\left(\frac{t - tp}{tn}\frac{\pi}{2}\right)\right] \quad \text{con } tp \le t \le tp + tn, \text{ donde t es el periodo fundamental.} \qquad (2.1)$$

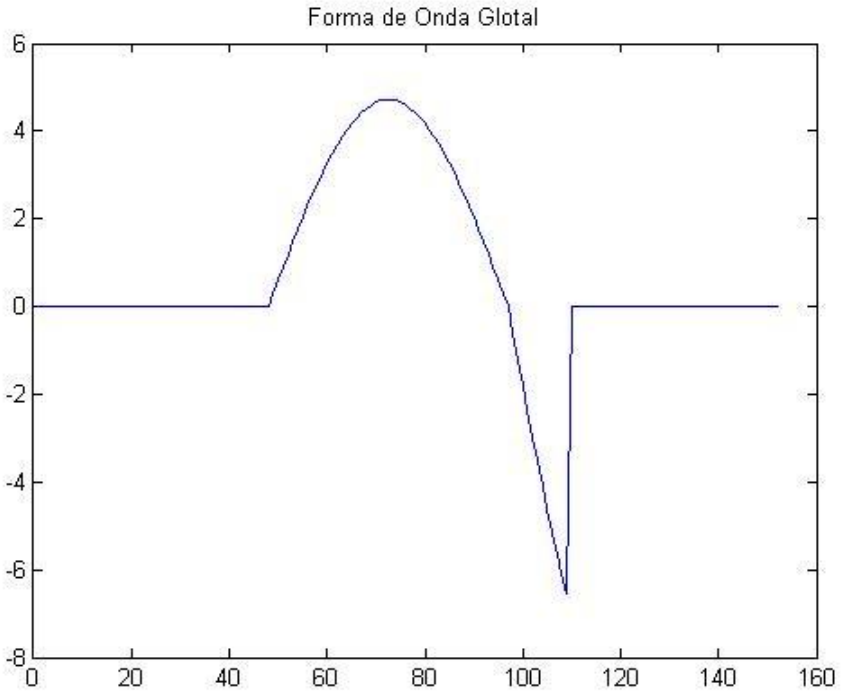


Figura 2.1.4. Pulso Glotal para la señal de excitación basado en Rosember.

Se generó un tren de impulsos de una glotis y se usó una derivada de un pulso glotal en el cual se excita el tren de pulsos con los siguientes parámetros: la perturbación, valor por defecto pert = 0, para pert = 0: ningún shimer, ningún jitter. Si pert= 1 se genera la vocal con jitter por el modelo S/A, si pert = 2, se genera la vocal con jitter por el modelo E/C y A/E.

2.2 Síntesis de vocal sostenida perturbada por jitter (modelos E/C y S/A)

Para modelar el tracto vocal para sintetizar una vocal 'A' se utilizó la función de trasferencia que se muestra en la Figura 2.2.

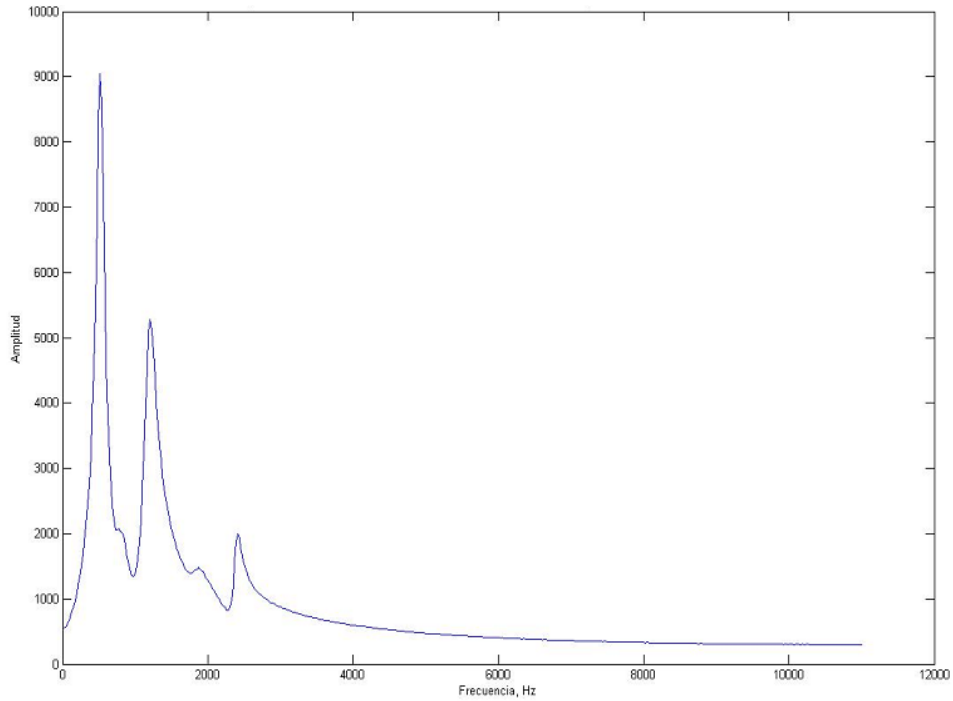


Figura 2.2. Función de trasferencia del tracto vocal utilizada para generar una vocal A.

Los modelos utilizados en el presente estudio son los modelos de S/A (jitter por posición) y E/C (jitter por duración), expuestos con anterioridad en la sección 1.4. Debe conocerse, que no se tiene por conocido ningún estudio anterior que evalúe simultáneamente ambos modelos. En la bibliografía que se encuentra localizada, respecto al tema, no se considera tan siquiera la posibilidad de la existencia de un modelo alternativo al asumido por los autores.

2.3 Base de datos de señales glotales reales

La realización de las pruebas para los modelos utilizados fueron tomadas de una base de datos, dicha base fue creada por voces de mujeres y hombres, está compuesta por 2366 frases fonéticas, que se toman del corpus la de la base de datos TIMIT [29]. El fondo acústico está formado sólo por el zumbido de la máquina de registro, el cual se encuentra a

2 m de la cabeza del sujeto y se enlaza por medio de un micrófono estéreo en el cual un canal se utiliza para grabar la voz y el otro la señal de electroglotografía.

Todas las grabaciones recogidas fueron supervisadas y llevadas a cabo en un lugar hermético sin perturbaciones ni influencia de ruidos externos, el ruido del fondo de la grabación portátil fue separado por una pared absorbente, el locutor y el supervisor se encontraban sentados en el cuarto magnetofónico donde el primero podía observar su propia pantalla. El locutor estaba provisto con el micrófono y con los electrodos que se encontraban a una distancia del micrófono de 1 a 2cm. Las frases fueron leídas, por lo que se garantiza la adecuada correspondencia entre el texto deseado y la grabación realizada. El software se implementó en una computadora portátil IBM, donde el micrófono empleado es modelo AKGHC577L y el auricular posee como características un rango de frecuencia de 80Hz a 20kHz, una sensibilidad de 35mV/ Pa(-29dBV), máximo SPL de 126dB/130dB (a 1%/3% THD), nivel de ruido equivalente de 22dB-A, relación señal/ruido(A-ponderada) de 72dB, impedancia menor igual a 200 ohms, impedancia de carga recomendada a mayor igual a 2000 ohmios y el consumo de corriente menor o igual a 2mA [30].

El dispositivo magnetofónico es un *Firewire* que graba la interface *Presonus Firebox* con las especificaciones técnicas de los signos grabados a señales electroglotografía que se registraron a una frecuencia de muestreo de 48 kHz, 16 bits de resolución, con el tipo de codificación PCM y el tipo de orden de bytes *Little Endian*; con dos canales, que registran en un archivo WAV estéreo. El canal izquierdo se utilizó para las señales de micrófono y señales electroglotográficas como archivos WAV de un solo canal. Para los datos de referencia del periodo fundamental, se utilizó un archivo de formato ASCll con la extensión *"f0"*. Este trabajo fue realizado en el *Institute of Broadband Communications at Graz University of Technology*. [31].

Estructura de la base de Dato.

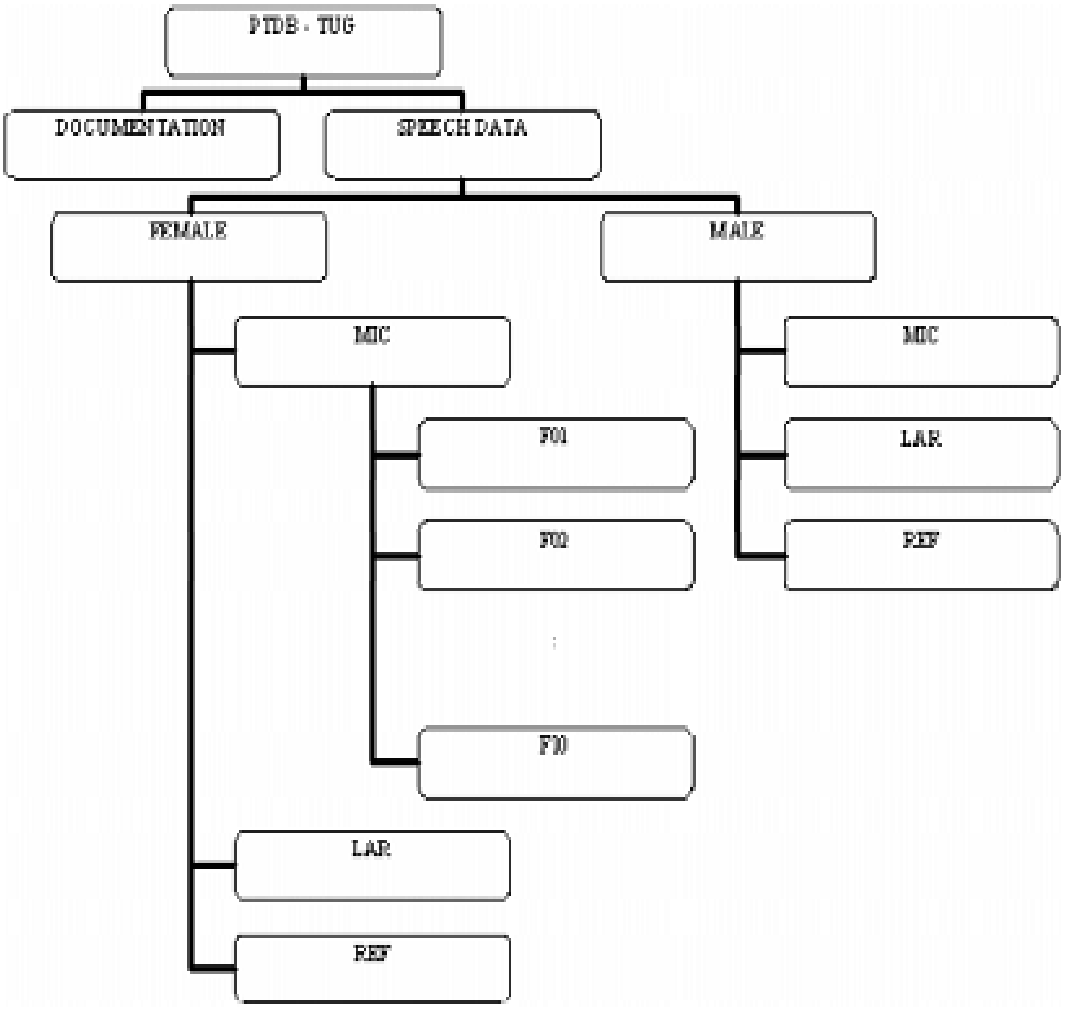


Figura 2.3. PTDB-TUG consiste en dos subdirectorios que contienen la documentación y los datos de voz. Los archivos de la señal del habla se dividen en datos femeninas y data masculino en primer lugar, y en señales de micrófono, señales electroglotografía y las señales de referencia en el segundo lugar. En cada uno de estos tres directorios se pueden encontrar carpetas etiquetados de acuerdo con los ID del habla que contienen los datos correspondientes.

MIC: señales grabadas por un micrófono

LAR: Señales grabados por el electroglotografía

REF: referencia extraída del pitch de trayectorias

2.4 Creación de un modelo del tracto vocal utilizando la Predicción Lineal

Desde que el término predicción lineal fue acuñado Wiener [32], esta técnica ha sido ampliamente empleada en un extenso rango de aplicaciones bajo distintas formulaciones. Utilizada por primera vez para el análisis y síntesis del habla, produjo gran impacto, en todos los aspectos del tratamiento del habla, dígase así: en estudios lingüísticos, en la logopedia, para determinar acertadamente ciertas patologías del habla, entre otras.

La técnica de predicción lineal, el cual se reconoce como LPC *(Linear Predictive Coding),* consiste en estimar el valor actual de una señal x(n) como una combinación lineal de las muestras anteriores. El valor estimado x(n) se escribe como

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k), \tag{2.2}$$

donde p es el orden de predicción y *ak* son los coeficientes de predicción. El problema básico de la predicción lineal consiste en determinar estos coeficientes *ak* de forma que la aproximación de x(n) sea suficientemente buena de acuerdo con algún criterio. El error entre la valor real x(n) y el valor estimado x^(n) se denomina error de predicción y viene dado por la expresión

$$e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^{p} a_k x(n-k). \tag{2.3}$$

A partir de esta expresión, puede considerarse el error de predicción como respuesta a x(n) de un sistema, que se denomina filtro de error de predicción, cuya función de transferencia es

$$A(z) = 1 + \sum_{k=1}^{p} \alpha_k \, z^{-k} \tag{2.4}$$

Además, a partir de (2.4) también puede escribirse

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n). \tag{2.5}$$

Por tanto, el modelo de predicción lineal de generación de señal puede representarse como

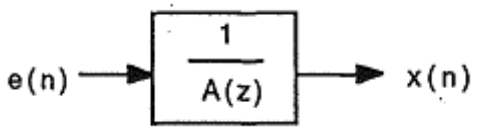


Figura 2.4. Modelo de generación de señal de predicción lineal.

En este trabajo el cálculo de los coeficientes de predicción lineal se realizó utilizando el método de la covarianza.

2.5 Descripción de los experimentos realizados.

Se realizaron tres experimentos utilizando señales simuladas. El **experimento uno** se realiza con señales simuladas sin perturbación, mientras que en los **experimentos dos y tres** se perturba la señal con 7 niveles de perturbación con jitter según los modelos de S/A y E/C. En ambos casos los niveles de perturbación comprenden un porciento de error de 3.4 % hasta un 23.8 % Para cada uno de los experimentos se realizan 300 repeticiones. A partir de estas señales se obtiene un modelo del tracto vocal para vocal mediante filtrado inverso, utilizando como coeficientes del filtro los coeficientes de predicción lineal

La validación se realiza calculando los errores cometidos en la detección del pulso glotal y comparándolos con la señal de pulso glotal generada sintéticamente, que fue realizada a cada una de las tramas de la señal sintética. , en todos los casos se evaluó la capacidad del modelo para recuperar la información de la fuente glotal a partir de obtener los instantes de cierre glotal y compararlos con las señales utilizadas en la generación de la voz sintética. La Figura 2.5 muestra un diagrama de flujo del procedimiento.

Se realizó además un experimento con señales reales de la base de datos referenciada en el epígrafe 2.3 donde para cada trama de la señal se conoce la frecuencia fundamental, el instante de cierre y el instante apertura glotal. El experimento se realiza tomando 10 señales de habla continua grabadas a igual número de locutores en idioma inglés, 5 de ellos mujeres y 5 de ellos hombres. En todos los casos la grabación realizada es una oración leída a un ritmo normal con una frecuencia de muestreo de 48 kHz. Se utilizan locutores de ambos sexos teniendo en cuenta las diferencias que existen en cuanto a *pitch* entre ambos. El experimento consiste al tener la voz grabada se le aplica un sistema de predicción lineal

al habla continua para obtener a partir del filtrado inverso, la señal de excitación glotal. Teniendo la señal de excitación glotal se calculan los instantes de cierre y apertura y se comparan con los instantes de cierre y apertura obtenidos a partir de la señal EGG de la base de datos.

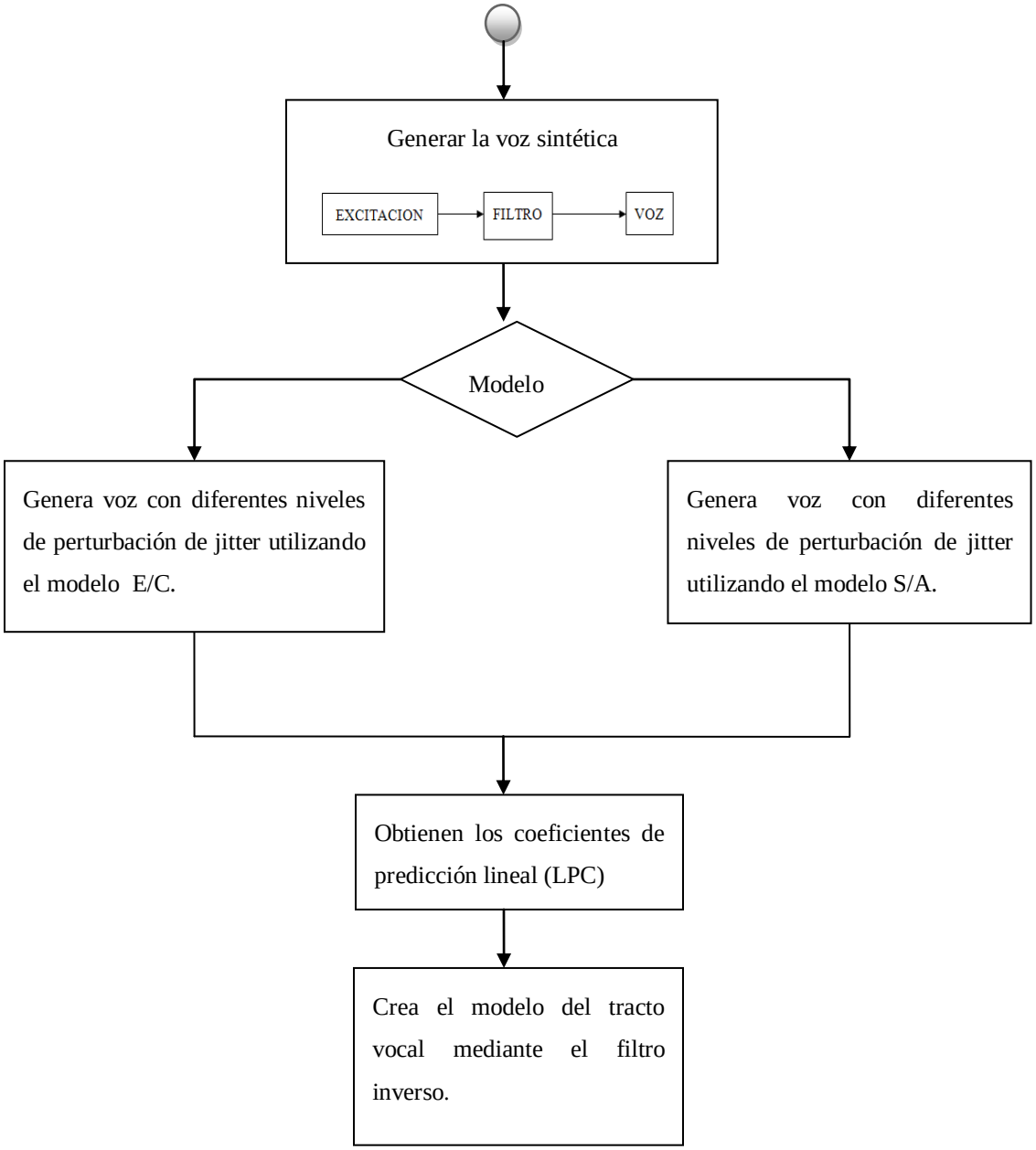


Figura 2.5. Diagrama de flujo del experimento.

2.6 Conclusión del Capítulo

En este capítulo se realizaron los experimentos para evaluar la influencia de dos modelos de variación del pulso glotal en función del jitter con la efectividad de medidas de perturbación de la periodicidad, a través de la base de datos de las señales glotales donde se empleó la predicción lineal.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En este capítulo se muestran los resultados de los experimentos realizados para validar las contribuciones propuestas en este trabajo y se comentan las implicaciones que pueden extraerse de estos resultados.

3.1 Evaluación utilizando señales sintéticas.

La evaluación se realizó generando señales que presentan jitter por posición y jitter por duración con 7 niveles de perturbación que implican la presencia de jitter desde un 3.4 % hasta un 23.8 %, en todos los casos se evaluó la capacidad del modelo para recuperar la información de la fuente glotal a partir de obtener los instantes de cierre glotal y comparalos con las señales utilizadas en la generación de la voz sintética.

3.1.1 Evaluación cuando no ocurre Jitter.

El resultado de la evaluación “sin Jitter” se muestra en la Figura 3.1.1, en la cual se puede apreciar la muestra del error en la obtención de los instantes de cierre glotal a partir de la señal glotal obtenida utilizando el filtrado inverso de la señal de voz sintética corespondiente a la vocal A sostenida; donde se observan las diferencias entre el instante glotal calculado y el real para señales sintéticas sin perturbaciones de jitter,el cual es prácticamente cero.

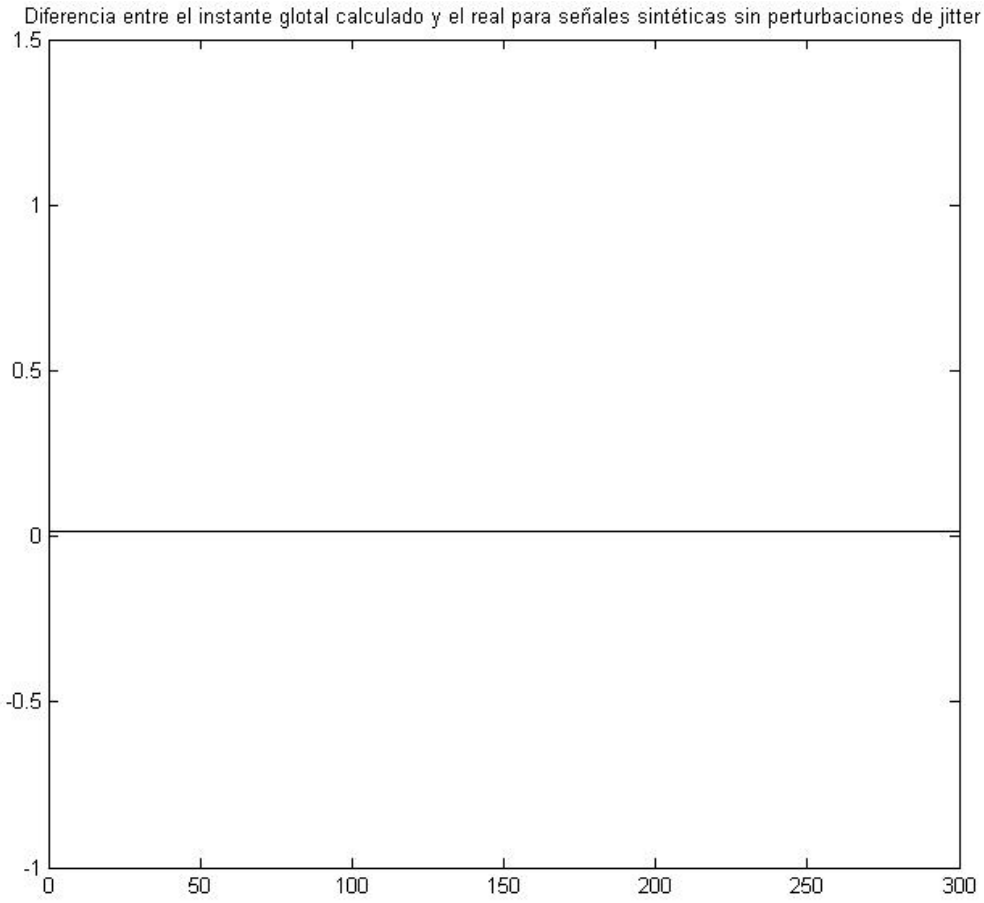


Figura 3.1.1. Error del instante de cierre glotal de la vocal sostenida.

3.1.2 Evaluación en presencia de Jitter por posición.

Se puede presenciar las diferencias entre el instante glotal calculado y real para señales sintéticas utlizando el modelo de jitter por posición S/A con siete niveles de perturbación, en el cual se realiza la evaluación de las señales en los errores en presencia de jitter y ver las diferencias entre estos niveles, que no varían unos de otros con valores muy parecidos desde 0.06 hasta 0.13 errores.

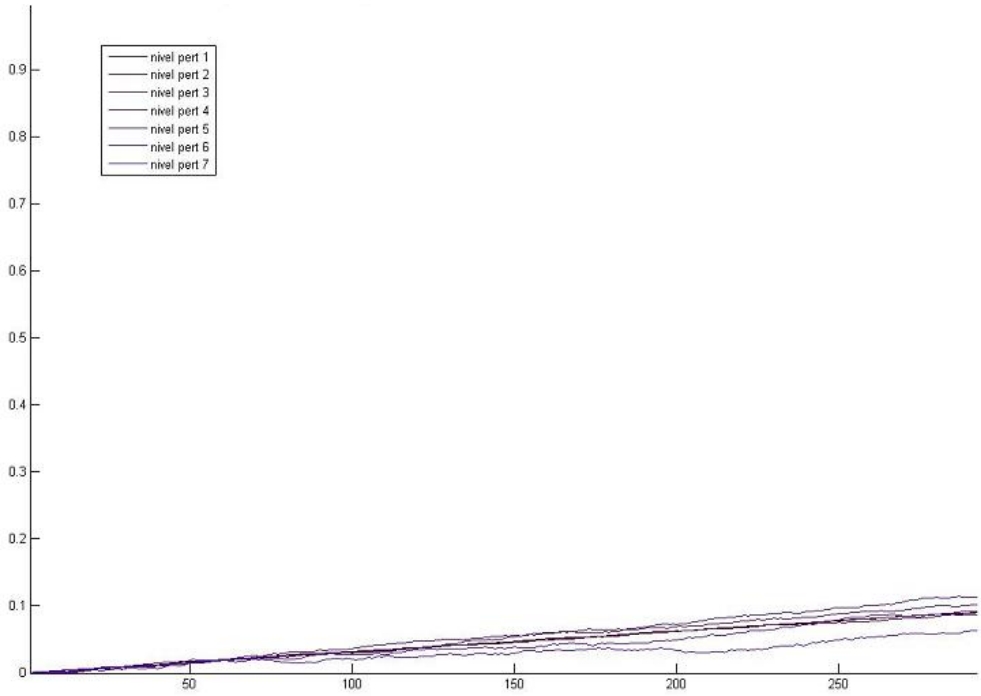


Figura 3.1.2 Error cometido para los 7 niveles de perturbación con jitter por posición.

3.1.3 Evaluación en presencia de jitter por duración.

Se evalúa el error de la señal en sus sietes niveles de jitter por duración y se puede presenciar las diferencias entre el instante glotal calculado y real para la señal sintética utlizada en el modelo de jitter por duración E/C y comparar el instante real y sintético, ver así como se va afectando la señal por cada nivel desde 4 hasta 6 errores.

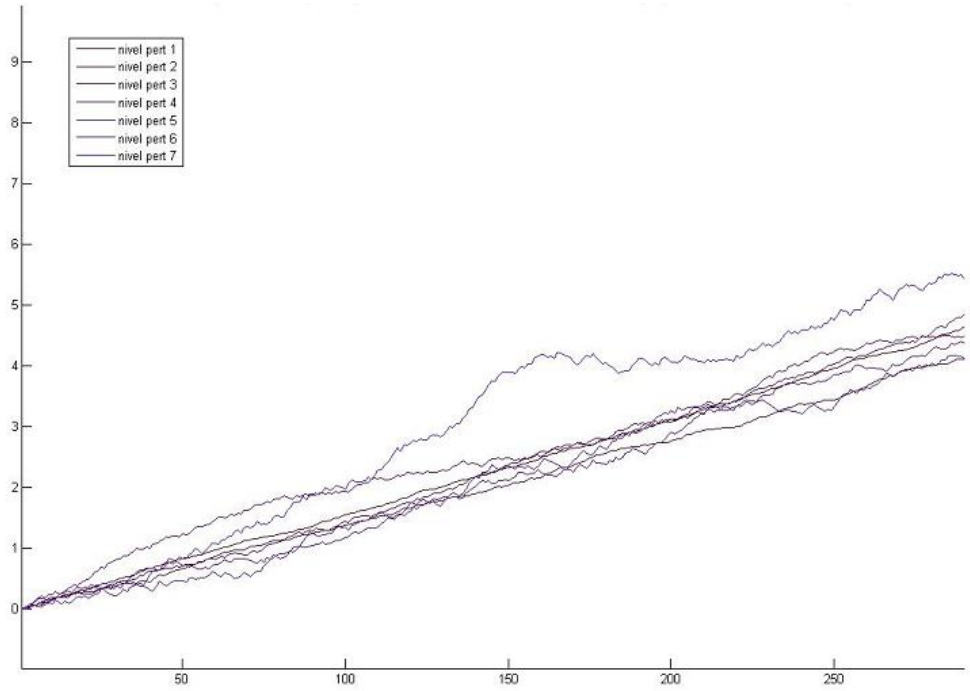


Figura 3.1.3. Error cometido para los 7 niveles de perturbación con jitter por duración.

3.1.4 Comparación de la influencia de los dos modelos de jitter tratados utilizando como línea base de la evaluación cuando no ocurre jitter.

Tabla 3.1.4. Resultado de la comparación de los modelos de jitter en cuanto a niveles de perturbación y las líneas de base de cuando no ocurre jitter.

Modelo de Jitter	Nivel de Perturbación						
	3.4	6.8	10.2	13.6	17.0	20.0	23.8
	Error						
Sin jitter	0.0029						
S/A	0.1167	0.1137	0.1184	0.1218	0.1138	0.1158	0.1216
E/C	0.8292	0.9065	1.2118	1.3535	1.6965	2.6718	2.9366

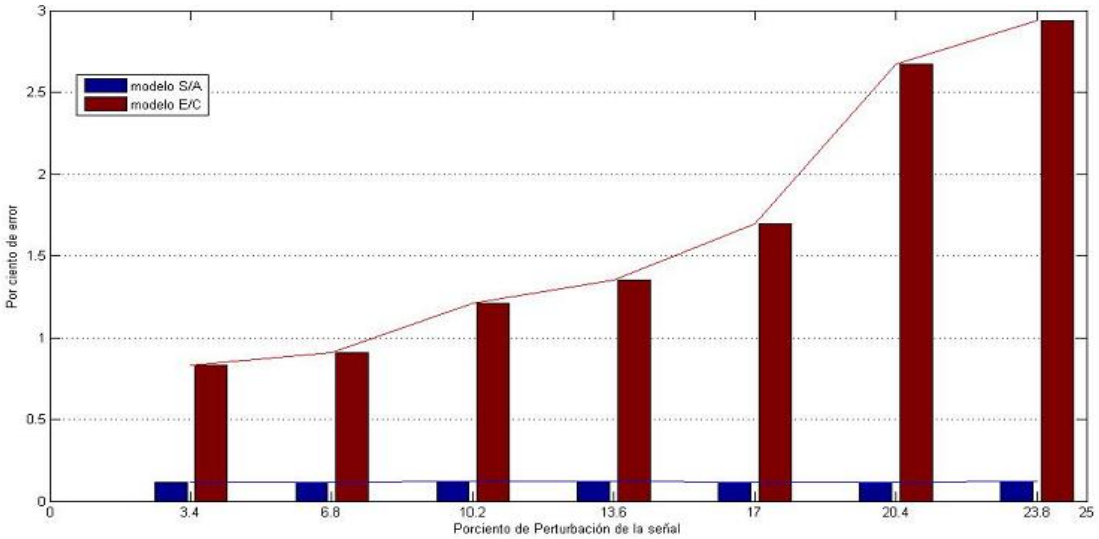


Figura 3.1.4 Resultados gráficos de la comparación de los modelos de jitter en cuanto a niveles de perturbación y las líneas de base de cuando no ocurre jitter.

En este resultado se puede observar la variación en el modelo de jitter por duración y posición (E/C y S/A) en cuanto a sus niveles de perturbación y el error promedio de los instantes glotales utilizando cada uno de estos modelos de jitter, aplicados a 100 muestras donde se ve la diferencia entre un modelo E/C y S/A, donde el primero aumenta considerablemente su error y el segundo se mantiene prácticamente lineal.

3.2 Evaluación utilizando voz grabada.

A modo de validación, se aplicó el procedimiento a un fragmento de señal de la base de datos, para obtener la función de trasferencia del tracto vocal y luego filtrar para obtener la señal de excitación glotal.

3.2.1 Obtención de la función de transferencia del tracto vocal.

Como resultado de la LPC por trama se obtienen los coeficientes del filtro de producción de voz con los que se calcula la función de transferencia del tracto vocal de una trama de voz continua, obteniendo así la Figura 3.2.1.

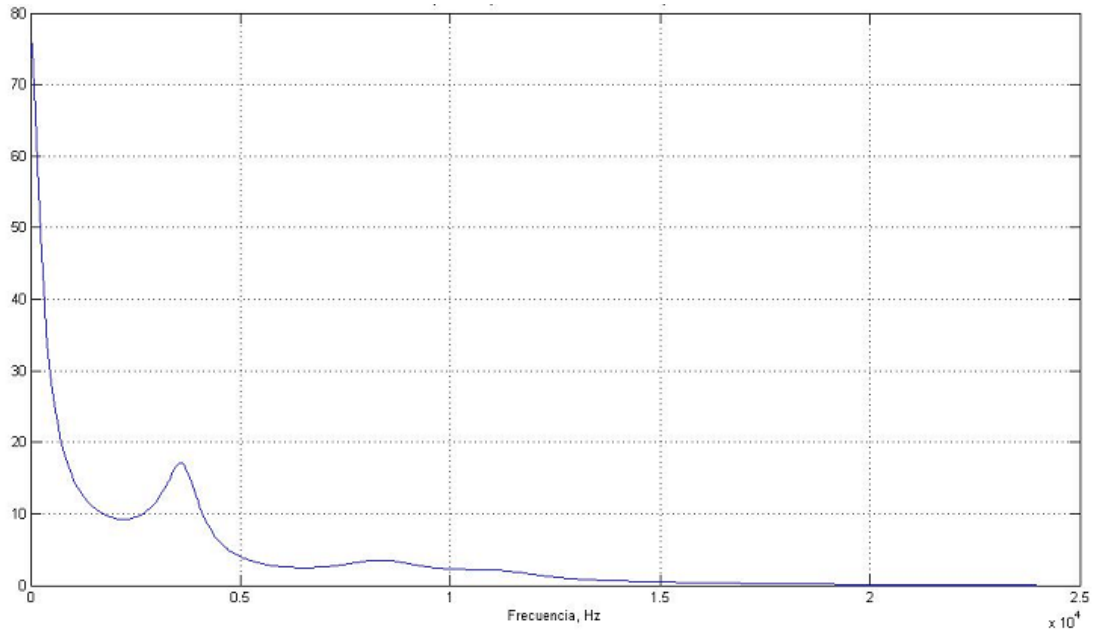


Figura 3.2.1. Resultado de la función de transferencia.

3.2.2 Evaluación en señales reales.

En este resultado se utilizó una frase de la base de datos en el cual se tomó un segmento de dicha frase de 10.5 segundos en donde se dividió en tramas de 20 ms y se le aplicó el LPC por tramas y a partir de ahí se calcula filtro inverso. El filtrado inverso se hace con enventanado y cada trama se filtra por el filtro inverso obteniéndose la forma de onda glotal que se muestra en la Figura 3.2.2.

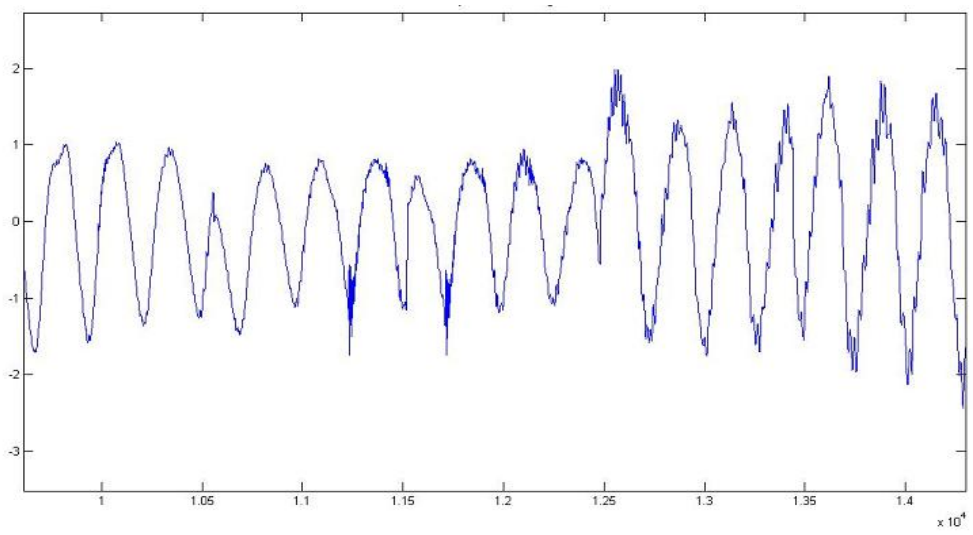


Figura 3.2.2. Resultado de la forma de Onda Glotal recuperada de un fragmento de voz continúa.

Al obtener esta forma de onda glotal se calculan los instantes de cierre y apertura de la glotis utilizando el algoritmo de análisis multiescala descrito en [33], obteniendo en un porciento de error que se puede apreciar en la Tabla 3.2.2.

Tabla 3.2.2. Resultado de la obtención de los porcentajes de los errores de instantes de cierre y apertura.

Alocución	% Error T_cierre	% Error T_ apertura
M1	7.9779	9.8123
M2	6.2888	6.1462
M3	16.1262	21.9376
M4	7.3908	8.1121
M5	7.7473	7.9692
F1	34.1343	41.2073
F2	11.9488	12.8938
F3	15.3874	17.0121
F4	10.7998	9.8296
F5	48.7262	53.6137

En general los errores cometidos para seguir los instantes de cierre fueron menores que los cometidos para seguir los instantes de apertura. Además de los errores reflejados en la tabla, también se cometen errores de inserción y rechazo de instantes de cierre y apertura. Las Figuras 3.2.3 y 3.2.4 muestran la apertura y cierre glotal.

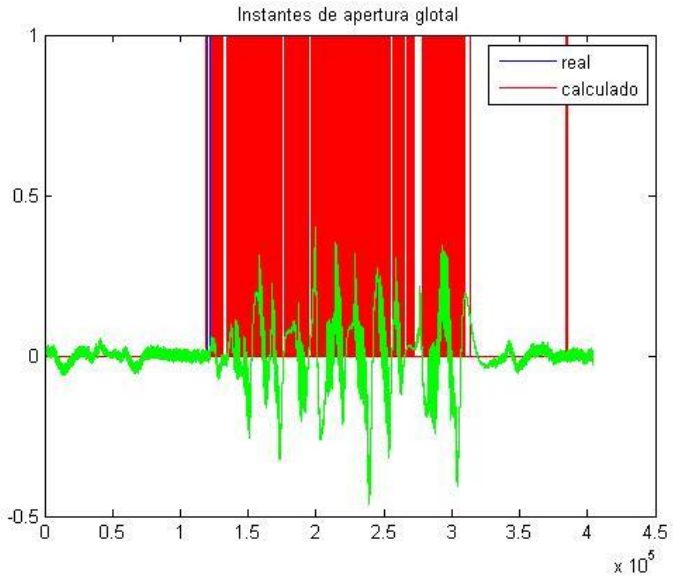


Figura 3.2.3 Resultado de los instantes de apertura glotal.

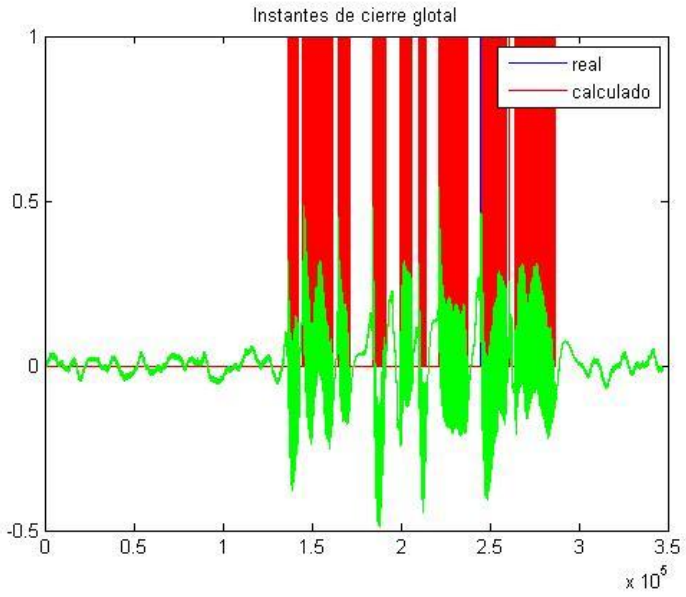


Figura 3.2.4 Resultado de los instantes del cierre glotal.

3.3 Conclusiones del Capítulo

En este capítulo se comparó el desempeño del modelo de tracto vocal sugerido frente ambos modelos de Jitter comprobándose que en los niveles de perturbación E/C tiene errores más altos que S/A.

Se comparó gráficamente el resultado de la variación en el modelo de jitter por duración y posición (E /C y S/A) en cuanto a sus niveles de perturbación y el error promedio de los instantes glotales utilizando cada uno de estos modelos de jitter, aplicados a 100 muestras donde se ve la diferencia entre un modelo E/C y S/A, donde el primero aumenta considerablemente su error y el segundo se mantiene prácticamente lineal.

Se obtuvo el resultado de aplicar el modelo a señales reales, donde se logró la función de transferencia del tracto vocal recuperado, así como los coeficientes de predicción lineal (LPC) de la voz continua. Los resultados preliminares sugieren que los instantes de cierre se siguen mejor que los instantes de apertura.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Durante la realización de este trabajo se obtuvieron resultados preliminares en la creación de un modelo del tracto vocal que sugiere las siguientes conclusiones:

1-El modelo de predicción lineal, empleado para modelar la función de transferencia del tracto vocal para obtener la señal de excitación glotal, es más sensible frente al modelo de jitter por duración que frente al modelo de jitter por posición.

2-En el caso del modelo de jitter por posición, los resultados preliminares sugieren que no se debe afectar el desempeño del modelo de predicción con el incremento de los niveles de perturbación. Este resultado es contradictorio, porque al aumentar los niveles de perturbación debería afectarse el modelo. En el caso del modelo de jitter por duración sí existe una relación entre los valores esperados y lo obtenido, deteriorándose sustancialmente el desempeño a medida que aumenta la perturbación.

3-El modelo de predicción lineal que se aplica en este trabajo puede mostrar buenos resultados en su uso en habla contínua, los resultados muestran que los instantes de cierre glotal se pueden describir mejor que los intantes de apertura.

Recomendaciones

Con la realización de este trabajo queda abierta la posibilidad de realizar trabajos futuros orientados fundamentalmente a:

1-Desarrollar un experimento que valore el desempeño del modelo en señales reales patológicas con el apoyo de especialistas.

2-Extender los experimentos a otros factores que afectan la periodicidad de la señal de voz como puede ser el shimmer.

REFERENCIAS BIBLIGRÁFICAS

[1] Laver, J. (1991). *The gift of speech*. Edinburgh, Edinburgh University Press.

[2] Hirano, M. (1981). *Clinical examination of voice*. Wien; New York: Springer-Verlag.

[3] Bodt, M.; Wuyts, F.; Van de Heynings, P. & Croux, C.T. (1997). *Test – Retest Study of the GRBAS Scale*: Influence of Experience and Professional Background on Perceptual Rating of Voice Quality. Journal of Voice, 11 (1), pp. 74 – 80.

[4] Dejonckere, P. H.; Bradley, P.; Clemente, P.; Cornut, G.; Crevier-Buchman, L.; Friedrich, G.; et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques-Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). European Archives of Oto-Rhino-Laryngology, 258 (2), pp. 77 – 82.

[5] Vasilakis, M.; Stylianou, Y. (2009). Voice Pathology Detection Basedeon Short-Term Jitter Estimations in Running Speech.

[6] Alexis Gonzales. (2011). Influencia de Modelos de Jitter en Medidas de Perturbación de la Periodicidad. Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI).

[7] Duque Sánchez, C.; Morales Pérez, M. (2007). Caracterización de voz empleando análisis tiempo-frecuencia aplicada al reconocimiento de emociones. Línea de Instrumentación y Control. Universidad Tecnológica de Pereira.

[8] Lara Peinado, A. J. (2006). Corrección Experimental de lesiones Iatrógenas de la cuerda vocal modelo experimental canino.

[9] Dimitar D. Deliyski. (1969–1972). "Acoustic model and evaluation of pathological voice production." Proceedings Eurospeech '93, Vol. 3.

[10] Agüero, P. D.; Tulli, J. C.; González, E. L.; Uriz A. J. y De la Cruz Arbizu, F. (2011). SAV: un sistema de análisis acústico para la evaluación de la voz.

[11] Maria Botero Tobon, L. (2010). Caracterización de los indicadores acústicos de la voz de los studios del programa licenciatura en música de la universidad de Caldas.

[12] Fourcin, A. (2000). "Voice Quality and Electrolaryngography" en R. D. Kent & M. J. Ball (eds) Voice Quality Measurement. Singular. San Diego.

[13] Cranen, B. & DE JONG, F. (2000). "Laryngostroboscopy" en R. D. Kent & M. J. Ball (eds) Voice Quality Measurement. Singular.

[14]Rosemberg, A. E. (1971). Effect of glottal pulse shape on the quality of natural vowels. Journal of the Acoustical Society of America, 84, 583 – 588.

[15] Fant, G., Liljencrants, J. & Lin, Q. (1985) A four parameter model of the glottal flow. Speech Transmission Laboratory Quarterly Progress Status Report No 4. Suecia, Royal Institute of Tecnology.

[16] Edward L.Riegelsberger, Ashok K. Krishnamurthy, (1993). Glottal source Estimation: Motheds of applying the LF model to inverse filtering Departamento of Electrica Engineering. The Ohio State University Columbus, OH 43210.

[17] Ken-Ichi Sakakibara, MorinosatoWakamiya, Hiroshi. (2005). A many-parameter model of laryngeal flow with ventricular resonance and supraglottal vibration. NTT Communication Science Laboratories, NTT Corporation, Atsugi-shi,243-0198, Japan, Imagawa Departmentof Otolaryngology, The University of Tokyo,113-0033.

[18] VuNgocTuan&Christophed'Alessandro, (1999). Robus Glottal Clourse Detection using The Wavelet Transform. LIMSI-CNRS. BP133, F91403Orsay, France.

[19]Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer" J. Acoust Soc. Am., 67 (3), pp. 979 – 990.

[20]Kadambe, S. and Boudreaux-Bartels, G. F. (1992). "Application of the wavelet transform for pitch detection of speech signals". IEEE TRANSACTIONS ON INFORMATION THEORY, 38 (2).

[21]Monzo, C.; Alías, F.; Iriondo, I.; Gonzalvo, X. y Planet, S. (2007). “Discriminating Expressive Speech Styles by Voice Quality Parameterization”. Saarbrücken, ICPhS, pp. 2081 – 2084.

[22]Titze, I. R. (1984). "Parametrizatio n of the glottal area, glottal flow, and vocal fold contact area." JASA 75: 570-580.

[23]Drew, M. (2003). "Modeling vocal fold moti on with a hydrodinamic semicontinuum model." JASA 114: 455-464.

[24]Story, B. (2002). "An overview of the physiology, physics and modeling of the sound sourse for vowels."

[25]Little, M. (2005). "A simple nonlinear model of vocal dynamics for synthesis and analysis." NOLISP-2005: 188-203.

[26]Veldhuis, R. (1998). "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation". JASA 103: 566-571.

[27]O'Leidhin, E. M. P. J. (2003). "Preliminary Glottal Source modeling for pathologic voices". Proc Maveba Conf., Florence.

[28]Ferrer, C. A. (2005). Cuantificación de parámetros de la voz para aplicaciones médicas. Centro de Estudios de Electronica y Tecnologias de la Informacion. Santa Clara, Universidad Central "Marta Abreu" de las Villas.

[29]Garofolo J.S.,Lamel L.F.(1990). Fisher W.M.,Fiscens J.G., Pallett D. S., Dahlgren N.L., Zue V.: readme.doc.TIMIT.

[30]D.Tallkin, (1995).”A robust algorithm for pitch traking (rapt),” Speech Coding and Synthesis, W.B. Kleijin and K .K Paliwal, Eds., pp. 495518.

[31]Gregor Pirker, (2011). Database for multi-PITCH Tracking, Craz, alpha: 1.0.

[32]Javier Hernando Pericas, F. (1993). Técnicas de procesado y representación de la señal de voz para el reconocimiento del habla en ambientes ruidosos. Barcelona.

[33]M. R. P. Thomas and P. A. Naylor. (2009). "The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals," IEEE Trans. Audio, Speech, Lang. Process., vol.17, no.8, pp.1557-1566.