UNIVERSIDAD DEL PAPALOAPAN.

Campus Loma Bonita

INGENIERÍA EN COMPUTACIÓN

MODELADO ESTADÍSTICO DEL ESTADO ANORMAL DE LA VOZ

TESIS

QUE PARA OBTENER EL TÍTULO DE:

INGENIERO EN COMPUTACIÓN

PRESENTA:

MARÍA EDITH VELÁZQUEZ VARGAS

DIRECTOR DE TESIS:

Dr. EDUARDO SÁNCHEZ SOTO

CO-DIRECTOR:

Dr. GIBRAN ETCHEVERRY DOGER

PROFESOR DE LA UNIVERSIDAD DE SONORA

LOMA BONITA, OAXACA, ENERO 2012

La Juventud

Se es joven cuando se ve la vida
como un deber y no como un placer,
cuando nunca se admite la obra acabada,
cumplida, cuando nunca se cree
estar ante algo perfecto.

Se es joven si se está lejos de la docilidad
y el servilismo, si se cree
en la solidaridad y en la fraternidad.

Se es joven cuando se quiere
transformar y no conservar;
cuando se tiene la voluntad de hacer
y no de poseer; cuando se sabe vivir
al día, para el mañana; cuando se ve
siempre hacia adelante.

Cuando la rebeldía frente
a lo indeseable no ha terminado.
Cuando se mantiene el anhelo
por el futuro y se cree todo lo posible.

Cuando todo esto se posee,
se pueden tener mil años y ser joven.

Jesús Reyes Heroles

Dedicado a mi hermano Francisco Velázquez Vargas

Agradecimientos

A mis padres **Andrés e Ignacia** , *por darme la vida, la oportunidad de estudiar y por confiar en mi. Por motivarme a salir adelante y hacer siempre lo correcto. Se que para ellos significó realizar muchos sacrificios pero siempre se mantuvieron constantes con su amor y sus deseos de ayudarme.*

A mis hermanos **Francisco, Estefana, Cristóbal, Elizabeth, Pablo, Andrés, Mariela, Ezequiel y Santiago** , *por su apoyo durante todo el camino recorrido hasta llegar a esta meta. Por siempre creer en mi y desearme lo mejor.*

A mi tía **Antonia** *por su apoyo y cariño.*

A **mi tía Josefina y sus hijas Virginia y Macrina** *por permitirme entrar en su familia y por cuidarme de mi cuando estaba lejos de casa.*

A **Israel** *por estar conmigo en mis momentos de desesperación y no permitir que desistiera. Por escucharme y ayudarme a recuperar la confianza en mi.*

A toda mi familia.

Al **Dr. Eduardo Sánchez Soto** *por darme la oportunidad de trabajar con el y compartir conmigo sus conocimientos para que pudiera llevar a cabo este proyecto.*

A mis revisores **Beatriz Carely Luna, Gibran Etcheverry Doger, Sergio Ivvan Valdez Pea, Eduardo Ortiz Hernndez** *por su entrega y dedicación para este proyecto .*

A los que me apoyaron para crear la base de datos utilizada en este trabajo, quienes me brindaron su tiempo y paciencia.

A PROMEP por el financiamiento proporcionado mediante el Proyecto PROMEP/103.5/10/5489.

A la **Universidad del Papaloapan** *por aceptarme como uno de sus estudiantes y darme la oportunidad de ser parte de ella.*

A todos **mis profesores** *quienes durante mi estancia en la universidad siempre estuvieron dispuestos a ayudarme.*

A todos mis amigos y compañeros que me acompañaron durante mis estudios.

A todos los Mexicanos, quienes hacen posible la educación pública en nuestro país.

Muchas gracias a todos!

Resumen

Dentro del presente trabajo, se abordará el estudio de la voz para la identificación de patologías, de manera a clasificarla como patológica o sana. Durante el análisis se utilizaron tres parámetros: los coeficientes cepstrum, el pitch y la pendiente del pitch. Esté último parámetro fue propuesto debido a cambios más abruptos de la magnitud del pitch con respecto al tiempo en las voces parológicas. El cálculo de estos parámetros se realizó con la ayuda de diversas herramientas, y a partir de estos parámetros se elaboraron modelos de mezcla de gaussianas (GMM). Al comparar los parámetros de las voces de los individuos que conformaron la población, con los modelos generados, se registraron similitudes entre las voces sanas, y de la misma forma entre las voces patológicas, lo que permitió una primera clasificación. Posteriormente, el modelo general de cada clase (voz patológica y sana) fue adaptado utilizando la voz de cada individuo, con lo cual, se obtuvieron los mejores resultados.

Palabras clave: *Detección de Patologías, Modelado Estadístico, Procesamiento de Señales.*

Abstract

In this work, we deal with a classification problem were speech is categorized into pathological and healthy. The used parameters are: cepstral coeficients, pitch and slop of the pitch. The last parameter was proposed in this work due to the abrupt changes of the pitch's slop presented on the pathological speech. Specialized software were used to compute the employed parameters, and those parameters were modeled using Gaussian Mixture Models (GMM). After compared the selected parameters, we observed similarities among the pathological speech and similarly for the healthy speech. This characteristics allowed a first classification. Subsequently, the obtained general models for each speaker and each class were adapted obtaining as a result better classification rates.

Key words: *Pathological Speech Recognition, Statistical Models, Signal Processing*

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

1.1. Antecedentes

El estudio de la voz se lleva a cabo desde mucho tiempo atrás; a lo largo de los años se han realizado estudios para determinar características de la voz de las personas que permitan extraer información útil para un fin determinado así como para encontrar la mejor manera de representarlos. De acuerdo con el fin que se persigue, podemos tener principalmente las siguientes áreas de aplicación: Reconocimiento de voz, Reconocimiento de locutor, Reconocimiento de Lenguajes, y de los más recientes, Reconocimiento de emociones y el Reconocimiento de patologías en la voz.

A continuación se explican brevemente cada una de estas áreas.

1.1.1. Reconocimiento de la voz.

Los primeros trabajos dentro del análisis de la voz, estaban centrados en el reconocimiento del habla y fueron llevados a cabo por AT&T [Álvarez2001]. Su principal objetivo era la creación de sistemas de reconocimiento que le permitieran automatizar los servicios ofrecidos por sus operadores.
Entre las aplicaciones más comunes del reconocimiento del habla se encuentran:

- Dictado automático.
- Control por Comandos.
- Sistemas diseñados para dar órdenes a la computadora.
- Telefonía.
- Sistemas Portátiles. Debido a su pequeño tamaño tienen muchas restricciones así que el habla es una solución natural para introducir datos en ellos.
- Sistemas Diseñados para Discapacitados.

1.1.2. Reconocimiento de locutor.

Conforme se realizaron avances en el estudio de la voz, la información que se obtenía a través de ella fue aplicandose en otras áreas, como por ejemplo, el reconocimiento de locutor.
El reconocimiento de locutor es un área de la inteligencia artificial y consiste en la identificación automática de una persona a través de su voz. El hecho de poder

identificar a una persona por medio de su voz radica en que las características de esta en cada persona son diferentes debido en su mayoría a la fisiología de su aparato de producción de la voz.

El reconocimiento de locutor puede clasificarse de diversas formas [Sánchez2009]:

- Según el texto pronunciado: Dependientes o independientes de texto
- Según la función que tiene: Identificar o verificar al locutor.

Un sistema de reconocimiento de locutor se conforma de las siguientes fases [Esteve2007].

- Adquisición de los Datos.
- Extracción de características.
- Evaluación/ Toma de decisiones.

Las técnicas utilizadas para el reconocimiento de locutor son variadas al igual que los parámetros utilizados [Sánchez2005].

Para el reconocimiento de locutor se ha hecho uso de los Modelos Ocultos de Markov (HMM) [Rabiner1989], y los modelos de Mezcla de Gausianas (GMM) [Reynolds2000]. Los primeros en para sistemas dependientes del texto y los segundos para sistemas independientes. En la mayoría de los casos los HMM y GMM han demostrado ser muy eficientes, aunque los modelos discriminantes (Maquinas de Soporte Vectorial, SVM) están ganando terreno actualmente [Sarria2009].

1.1.3. Reconocimiento de lenguajes.

El reconocimiento de lenguajes es otro de los campos en los que se aplica el análisis de la voz. Desde hace unos años se han realizado estudios en esta rama, sin embargo, su desarrollo se ha incrementado recientemente, debido a las necesidades actuales, y con el uso diario de computadoras [Rouas2005]. El amplio uso de estas maquinas ha creado la necesidad de obtener herramientas de interacción que permitan una buena comunicación. El reconocimiento de lenguajes es una de las herramientas que está teniendo un gran auge actualmente.

Lo anterior deriva del hecho del alto intercambio plurilingue en la sociedad actual, por ello las interacciones humano-computadora necesitan ser de la misma manera: plurilingues.

1.1.4. Reconocimiento de emociones.

Al igual que el mensaje y la información correspondiente a la identidad la voz conlleva impregnada las emociones (miedo, alegría, tristeza, etc.) del locutor [Calvel2007]. Su detección a través de la voz actualmente encuentra aplicaciones en materia de vigilancia entre otras. Detectar, a través de una voz solamente, que una persona está en posible peligro y con miedo puede ser de gran útilidad.

Si bien el tipo de modelo selecionado en esta aplicación es importante, la extracción de características lo es mucho más. El problema principal al que se enfrentan los investigadores en esta área es la disponibilidad de bases de datos. Es extremadamente difícil gravar la voz de una persona con algunos de los tipos

de emociones, por ejemplo el miedo.

1.1.5. Reconocimiento de patologías de la voz.

En la actualidad, los estudios realizados en el campo del análisis acústico de la voz, se centran en el reconocimiento de patologías de la voz. En algunos trabajos, se han utilizado los Modelos de Ocultos de Markov para seleccionar las características que ayuden a identificar patologías en la voz [Álvarez2005]. Estas patologías son, frecuentemente, aquellas relacionadas directamente con el conducto vocal, tales como la disfonía, laringitis, o enfermedades degenerativas como el Parkinson [Rigaldie2004, Kapoor2011].

Dentro del presente trabajo, se abordará el estudio de otras características que ayuden a la identificación de patologías en la voz.

1.2. Planteamiento del Problema

Por medio de los trabajos realizados en el análisis de la voz para el reconocimiento del locutor, sabemos que dependiendo de las características que tenga el aparato fonatorio de cada persona se obtendrá un sonido que conlleva las características de tal persona. Por lo tanto, con un análisis acústico podemos identificar a un individuo. Durante el análisis acústico se deben tener en cuenta algunos factores que influyen en el reconocimiento del locutor [Sanchez2009], tales como:

- La calidad de la señal. Sistema de grabación, acústica del lugar donde se realiza la grabación, tipo de micrófono, ruido en el ambiente.

- Variaciones con el tiempo a largo plazo. Estado de ánimo, salud.
- Cantidad y variedad del material hablado.
- Motivación / Cooperación del Locutor. Locutores que cambian su forma de hablar imitando voces o falseando las suyas.

La segunda observación es la que más nos interesa en este estudio, ya que cuando una persona está enferma, para ciertos padecimientos, el tono de su voz cambia con respecto a la que normalmente percibimos. El malestar causado por algún padecimiento afecta nuestro aparato fonatorio, impidiéndole producir el sonido como normalmente se hace y como resultado la voz cambia.

Estas variaciones, de ser medibles dependiendo de la enfermedad, nos permitirá determinar las características de la voz que han sido modificadas (frecuencia, ritmo, energía).

Una vez que se hayan definido estas características, se podrán medir, comparar y agrupar dependiendo de la naturaleza y del origen, hasta llegar a un modelo. Con este modelo se pueden realizar estudios que permitan identificar alguna enfermedad presente en una persona por medio de su voz.

Como la voz puede estudiarse sin la necesidad de que una persona este en el lugar donde se lleva a cabo el estudio, podemos crear un sistema que nos permitan obtener la señal de voz a distancia, procesarla e identificar posibles enfermedades en ella.

Un sistema de este tipo ayudaría a muchas personas que viven lejos de un centro de atención médica. Esto con el objetivo de que reciban una atención y puedan prevenir posibles complicaciones.

1.3. Objetivos

1.3.1. Objetivo General.

Análisis de la voz para determinar características que nos ayuden a modelar de manera eficiente el estado anormal de la voz.

1.3.2. Objetivos Particulares.

A continuación se describen los objetivos específicos que nos permitirán, una vez alcanzados, lograr el objetivo general.

- Colecta de muestras de voz patológica.
- Análisis de la voz para determinar características que denoten la existencia de alguna anomalía (Energía, Frecuencia fundamental, Pitch, Ritmo, Características Frecuenciales).
- Implementación del Modelo Estadístico que se adapte a las características seleccionadas.
- Pruebas de desempeño del sistema.

1.4. Hipótesis

La voz de una persona se define de acuerdo con las características con las que cuentan los órganos de su aparato fonatorio; debido a esto, si algunos de estos componentes sufren una modificación, la voz también cambia. Por lo tanto cuando una persona padece alguna enfermedad que afecta algunos de los órganos del aparato fonatorio, estos se reflejaran en modificaciones de la voz producida.

1.5. Justificación

En el proceso de producción de la voz, los articuladores en sí mismos y su posición definen las características de esta, por lo que cuando existe un problema en el conducto vocal la voz cambia y existe por ende una variación acústica de la voz. Si las características y la posición de los articuladores no es la que normalmente se tiene, esto da como resultado una señal de voz diferente a la normal. Este cambio en la posición de los articuladores puede ser generado por algún padecimiento, el cual impide a alguno de los articuladores funcionar de manera normal. Por ejemplo, las cuerdas vocales podrían estar afectadas por alguna anomalía, con la cual, y debido al malestar ocasionado por la anomalía, el locutor presenta problemas al generar la voz. Este escenario es común cuando se realizan exploraciones de la laringe en busca de alguna patología; los pacientes no pueden articular bien la voz porque el padecimiento le impide abrir bien la boca, o por tener obstruido el conducto nasal. En este ámbito, el análisis acústico de la voz, permite una exploración menos dolorosa en los pacientes afectados, ya que

la mayoría de las técnicas de exploración son invasivas. Además, los pacientes no tendrían que desplazarse hasta el centro de salud para realizar un estudio, basta con que hable por teléfono para analizar su voz.

Es importante acotar que el médico no puede ser sustituido. El análisis acústico solo proporcionará índices que ayudaran a tomar una decisión médica.

1.6. Alcances de la tesis

- Las características seleccionadas permitirán comparar la voz de una persona cuando esté sana y cuando está enferma, de manera que se puedan observar los cambios generados en la voz en ambos estados. Esto ayudará en la detección de patologías que estén relacionadas con el conducto vocal.

Capítulo 2

Análisis de la voz

La voz humana es una fuente muy importante de información. Esta, conlleva tanto información del mensaje que se quiere transmitir como información de la persona misma. Aunque el mecanismo de producción de la voz es el mismo para todos, cada uno de nosotros tenemos rasgos diferentes, con lo que la voz producida también lo es. Por lo tanto, y debido a estos rasgos se puede identificar a una persona, ya que cada una tiene un timbre de voz diferente.

La voz también lleva información del estado de ánimo e inclusive del estado de salud de una persona. Para cierto tipo de enfermedades, las características de la voz, tales como el ritmo, la energía, etc., son diferentes cuando una persona está enferma y cuando está sana.

Para entender mejor estos fenómenos comenzaremos por describir como se produce la voz.

2.1. Modelo de Producción de la voz

La producción de la voz humana se realiza por medio del aparato fonatorio, el cual está compuesto principalmente por los siguientes elementos [Rabiner2010]:

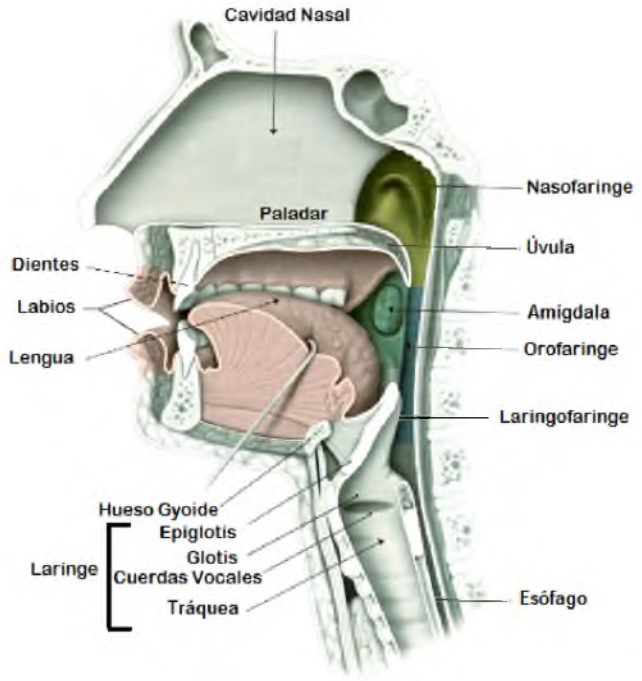

Figura 2.1: Aparato Fonatorio Humano

- **Pulmones** Fuente de energía, en la forma de un flujo de aire. Están situados dentro de la caja torácica, protegidos por las costillas. Son huecos y están cubiertos por una doble membrana lubricada llamada pleura. Están separados el uno del otro por el mediastino.

- **Cavidad Nasal** Es la parte interna de la nariz, comienza en el velo o paladar blando y termina en el nostril. Posee una forma rectangular con 4

paredes y 3 orificios de entrada y salida. Esta cubierta por epitelio respiratorio con cilios que permiten el barrido del moco producido por las glándulas mucosas.

- **Cavidad Bucal** - Se inicia en la glotis (cuerdas vocales) y termina en los labios. Consiste en la faringe (la conexión del esófago hasta la boca) y la boca en sí misma (la cavidad bucal). El área transversal que está determinada por posiciones de la lengua, labios, mandíbula y el velo, varía de 0 (cierre completo) a 20 cm^2.

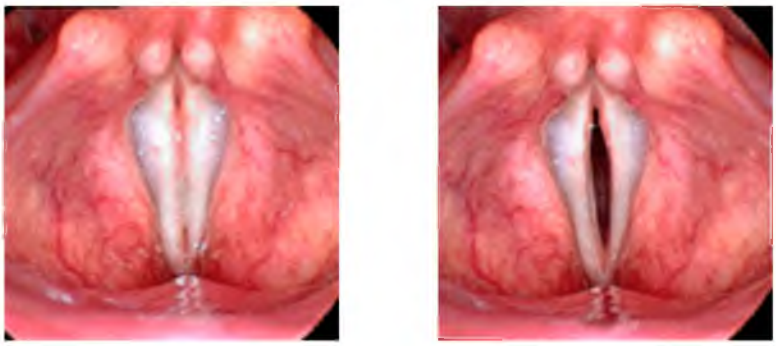

Figura 2.2: Cuerdas Vocales

Las cuerdas vocales son, en realidad, dos membranas dentro de la laringe orientadas de adelante hacia atrás. Por adelante se unen en el cartílago tiroides. Por detrás, cada una está sujeta a uno de los dos cartílagos aritenoides, los cuales pueden separarse voluntariamente por medio de músculos. La abertura entre ambas cuerdas se denomina glotis.

El mecanismo de producción de la voz, se realiza de la siguiente manera: El aire entra en los pulmones a través de la respiración normal. Cuando el aire es expulsado de los pulmones, a través de la tráquea, las cuerdas vocales dentro de

la laringe vibran debido al flujo del aire. Cuando la glotis comienza a cerrarse, el aire que la atraviesa, proveniente de los pulmones, experimenta una turbulencia emitiéndose un ruido de origen aerodinámico conocido como aspiración.

Al cerrarse más, las cuerdas vocales comienzan a vibrar a modo de lenguetas, produciéndose un sonido tonal. La frecuencia de este sonido depende de varios factores: del tamaño y la masa de las cuerdas vocales, de la tensión que se les aplique y de la velocidad del flujo del aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, lo cual explica porqu en los varones, cuya glotis es en promedio mayor que la de las mujeres, la voz es en general más grave. A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. El aire es cortado en pulsos cuasi-periódicos que son modulados en frecuencia (en forma de espectro), de paso a través de la faringe, la cavidad bucal y posiblemente, la cavidad nasal. Las posiciones de los diferentes articuladores (mandíbula, lengua, velo, labios, boca, etc.) determinan el sonido que se produce.

2.1.1. Clasificación de los Sonidos

Los sonidos emitidos por el aparato fonatorio pueden clasificarse de acuerdo con diversos criterios, estos criterios son [Miyara2004]:

a) Según su carácter vocálico o consonántico. Las vocales son los sonidos emitidos por la sola vibración de las cuerdas vocales sin ningún obstáculo o constricción entre la laringe y las aberturas oral y nasal. Las consonantes, por el contrario, se emiten interponiendo algún obstáculo formado por los elementos articulatorios.

b) Según su oralidad o nasalidad. Los fonemas en los que el aire pasa por la cavidad nasal se denominan nasales, en tanto que aquéllos en los que sale por la boca se denominan orales.

c) Según su carácter tonal (sonoro) o no tonal (sordo). Los fonemas en los que participa la vibración de las cuerdas vocales se denominan tonales o, también, sonoros. Aquellos fonemas producidos sin vibraciones glotales se denominan sordos.

d) Según el lugar y el modo de articulación (Consonantes). La articulación es el proceso mediante el cual alguna parte del aparato fonatorio interpone un obstáculo para la circulación del flujo de aire. Las características de la articulación permitirán clasificar las consonantes. Los órganos articulatorios son los labios, dientes, las diferentes partes del paladar (alveolo, paladar duro, paladar blando o velo), la lengua y la glotis. Salvo la glotis, que puede articular por sí misma, el resto de los órganos articula por oposición al otro.

e) Según el lugar o punto de articulación se tienen fonemas: Glotales (articulación en la propia glotis), Velares (oposición de la lengua con el paladar duro), Alveolares (oposición de la punta de la lengua con la región alveolar), entre otros. A su vez, para cada punto de articulación, esta puede efectuarse de diferentes modos, dando lugar a fonemas: Oclusivos (la salida del aire se cierra momentáneamente por completo), Laterales (la lengua obstruye el centro de la boca y el aire sale por los lados), Aproximantes (la obstrucción muy estrecha que no llega a producir turbulencia.) entre otras.

f) Según la posición de los órganos articulatorios (Vocales). En el caso de las vocales, la articulación consiste en la modificación de la acción filtrante de los diversos resonadores, lo cual depende de las posiciones de la lengua, de la

mandíbula inferior, de los labios y del paladar blando.

g) Según la duración. La duración de los sonidos, especialmente las vocales, tienen importancia en el plano expresivo, a través de la agogía, es decir, el énfasis o acentuación a través de la duración. En inglés, la duración de una vocal puede cambiar completamente el significado de la palabra que la contiene.

Esta clasificación nos será de utilidad al analizar la voz, ya que hay deficiencias que afectan a uno u otro sonido.

2.2. Características de la voz.

Una definición exacta de la voz normal no existe como tal, ya que la voz de una persona difiere de otra [Godino2007]. Una alternativa podría ser la cuantificación de la calidad de la voz, de manera que se tenga una referencia cuando se hable de voz normal.

Las cualidades más evaluadas en la voz son las siguientes:

- El timbre debe ser agradable.
- El tono debe ser adecuado para la edad y sexo del individuo.
- El volumen debe ser apropiado.
- La flexibilidad debe ser la adecuada.

2.3. Características de la voz patológica.

De acuerdo a resultados de estudios realizados en la voz, las características frecuentes de la voz patológica son las siguientes [Godino2007]:

- Aumento de las perturbaciones de la voz, en periodo y amplitud (jitter y shimmer).
- Presencia de ruido en el espectrograma.
- Disminución de los armónicos en el espectrograma.
- Presencia de subarmonicos.
- Ruido en alta frecuencia.
- Interrupciones o rupturas de la voz durante la fonación.
- Alteraciones morfológicas en los pulsos glóticos.
- Disminución del rango de fonación y/o rango dinámico.
- Aparición de componentes moduladoras en frecuencia y/o amplitud.

Dependiendo de la enfermedad con la que se esté trabajando será la característica que se tendrá, por ejemplo para las personas con Parkinson el ritmo disminuye, en cambio, cuando una persona padece de laringitis, tendrá otras características tal como interrupciones o rupturas de la voz durante la fonación.

Entonces, para identificar las enfermedades, primero se deben indentificar las características que la distinguen, es decir, que características de la voz son las

afectadas. En este proyecto se analizan las características de la voz de personas enfermas de gripa, se espera que más adelante se puedan definir las características de otras enfermedades.

2.4. Detección de Patologías de la voz

2.4.1. Estado del Arte

En estudios recientes y gracias a ayuda de los avances tecnológicos obtenidos con las computadoras, se han podido estudiar de manera más amplia los procesos de producción de la voz para aplicar estos conocimientos en otras áreas; un ejemplo claro es el caso del área médica con la detección de patologías por medio de la voz.

Aunque las patologías identificadas no son muy variadas se han tenido buenos resultados. En la mayoría de los casos la patología que se maneja es el Parkinson [Rigaldie2004]. De la misma forma, las locuciones son por lo general muy cortas, solo se ha utilizado el fonema $|a|$ [Rigaldie2004, Kapoor2011], principalmente porque es un sonido sonoro.

En un estudio reciente realizado por investigadores de la Universidad de Carabobo Venezuela [DelPino2004, DelPino2008], se utilizaron como parámetros característicos para clasificar las voces como sanas o patológicas, el pitch, formantes, energía, jitter y shimmer, además de tres parámetros de calidad en el dominio de la frecuencia: relación valor pico-valor medio en frecuencia (PMR), relación señal

a ruido - frecuencias medias (SNRM), y relación señal a ruido a frecuencias altas (SNRH). La patología identificada fue el Parkinson y para la adquisición de la señal de voz, los pacientes emitieron el fonema $|a|$ de manera sostenida.

Un estudio similar realizado en la india, utiliza los coeficientes cepstrales en la escala MEL y utilizaron la cuantificación vectorial para clasificar las voces de personas enfermas de Parkinson y personas sanas [Kapoor2011].

En otro estudio donde se estudia el Parkinson, se utilizó el sistema INTSYNT (International Transcription with a limited set of abstract tonal symbols) para estudiar los distintos niveles que tiene la enfermedad [Rigaldie2004].

En nuestro proyecto buscamos obtener parámetros que nos permitan clasificar las voces como sanas o patológicas; donde la enfermedad afecte directamente al conducto vocal. Todo esto con el objetivo de crear un sistema completo y general de identificación de patologías. Para lograr lo anterior se utiliza un texto de aproximadamente 1000 palabras para la señal grabada, a diferencia de los estudios anteriores que han hecho uso solo de un fonema.

Mediante el sistema de comunicaciones, se busca obtener un sistema médico que proporcione atención a las personas que estén alejadas de los centros de salud. En los países industrializados, donde la mayoría de la gente mayor vive sola, un sistema de prevención de este tipo sería de gran utilidad.

Sistemas de este tipo se están desarrollando ya, por ejemplo tenemos el proyec-

to Apollo Hospitals [Apollo], fundado en 1983 por el Dr. Prathap C. Reddy para dar atención a las personas en la India, quienes no tienen acceso a una atención médica.

También esta el proyecto MyGlucoHealth [GlucoHealth], el cual monitorea los niveles de glucosa de los pacientes a distancia, mediante el envío de la información por bluetooth.

Estos proyectos son posibles gracias a los avances que tenemos en los sistemas de comunicación, pero es necesario su estudio para adecuar su función a otras aplicaciones, como las que realizan en estos proyectos y en países no industrializados.

2.4.2. Patologías identificadas

En nuestro proyecto se utilizaron individuos enfermos de gripa, pues es una enfermedad que afecta principalmente al conducto vocal y los cambios producidos por dicha enfermedad son más comunes (la colecta de la base de datos es más fácil).

Los estudios que se lleven a cabo durante el proyecto con la ayuda de esta enfermedad, podrán permitir realizar estudios más adelante utilizando otras enfermedades y no solo la gripa.

Capítulo 3

Modelado Estadístico

Un modelo estadístico es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales, y en la regresión, para indicar los diferentes factores que modifican la variable de respuesta.

Se pueden distinguir dos categorías: discriminativos y no discriminativos. Los clasificadores discriminativos solo modelan los límites entre las clases para separar una clase de otra. En cambio, los clasificadores generativos construyen un modelo de la distribución de las clases y a partir de ella pueden generar datos artificiales.

En la categoría de los clasificadores discriminativos se encuentran las Máquinas de Soporte Vectorial (SVM) y el análisis discriminante.

La teoría de las **máquinas de soporte vectorial** fue desarrollada inicialmente por V. Vapnik a principios de los años 80 [González1993].

Para dar una definición de un **SVM** suponemos que tenemos dos clases, mostradas en la figura 3.1, las cuales tienen al menos un hiperplano que las separa. Las SVMs buscan entre todos los hiperplanos separadores, aquel que maximice la distancia de separación entre las dos clases posibles.

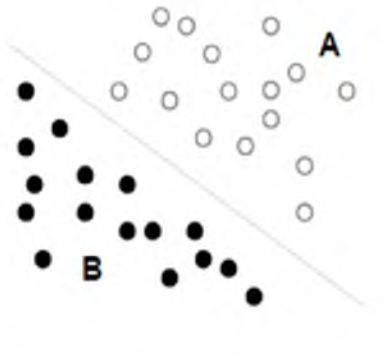

Figura 3.1: Conjunto e hiperplano separador en R^d . Los puntos huecos representan la clase 1 y los restantes la clase 2..

La solución para el caso de dimensión dos se puede interpretar gráficamente a partir de la figura 3.2. Con la ayuda de los vectores de soporte se definen las hiperplanos que dividen las dos clases. A la vista de esta figura, podemos darnos

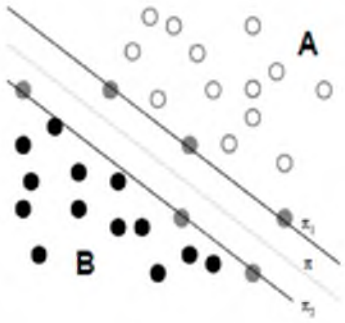

Figura 3.2: Hiperplanos paralelos y vectores de soporte en R^2.

cuenta que si se añade una nueva clase que se encuentre entre los dos hiperplanos, la solución cambia totalmente. Razón por la cual no se utilizan las máquinas de soporte en nuestro caso de estudio, ya que si tuviéramos que agregar un nuevo individuo al modelo tendríamos que volver a calcular los parámetros de toda la mezcla, en cambio con un GMM, solo se realiza una adaptación del modelo.

El objetivo del **análisis Discriminante** es proporcionar una regla discriminante que permita asignar un nuevo individuo u objeto a una de varias poblaciones, clases o grupos previamente identificados [Galbiati].

La regla se obtiene a partir de una muestra, consistente en un conjunto de observaciones multivariadas, en las que una variable es la población a la que pertenece cada observación. Existen varios métodos para obtener la regla discriminante, entre las más conocidas están el modelo de máxima verosimilitud clásico y el procedimiento discriminante canónico.

Para obtener una regla discriminante se dispone de una matriz de datos X, llamada muestra de aprendizaje. Una de sus variables (columnas) es un factor, que indica la población a la que pertenece cada una de las observaciones. Esta se considera como la variable dependiente. Las demás son consideradas como variables independientes.

Dentro de los clasificadores generativos tenemos los Modelos de Mezclas de Gausianas (Gaussian Mixture Models, GMM), los cuales se explican a continuación.

3.1. Modelos Generativos

A continuación se explican las características de los dos modelos generativos más comunes dentro de la literatura y el procesamiento del habla, los GMM y los HMM; los primeros para casos estáticos y los segundos para sistemas dinámicos.

3.1.1. Modelos de Mezcla de Gaussianas

Los modelos de mezcla de Gausianas (GMM) fueron aplicados a la verificación de locutor por primera vez en una serie de artículos publicados por Reynolds en 1990. A través de los años se han convertido en una aplicación dominante para el modelado en aplicaciones de reconocimiento de locutor.

Un GMM es un modelo de distribución de probabilidad de mezclas finitas. [Carvajal2010] Es utilizado para identificaciones de locutor independientes del contexto, ya que no le interesa lo que la persona está diciendo, únicamente extrae información espectral de la señal de voz [Reynolds1995]. La función de densidad de probabilidad de un GMM se puede definir como una suma ponderada de M densidades [Reynolds1995], como se muestra en la figura 3.3 y está dada por la siguiente ecuación:

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x})$$

Donde $\vec{x}$ es un vector de D-dimensiones aleatorias, $b_i(\vec{x}), i = 1, ...M$, son las componentes de densidad y $p_i, i = 1, ..., M$ son las medias de la mezcla. Cada componente de densidad es una función gausiana D-variada de la forma:

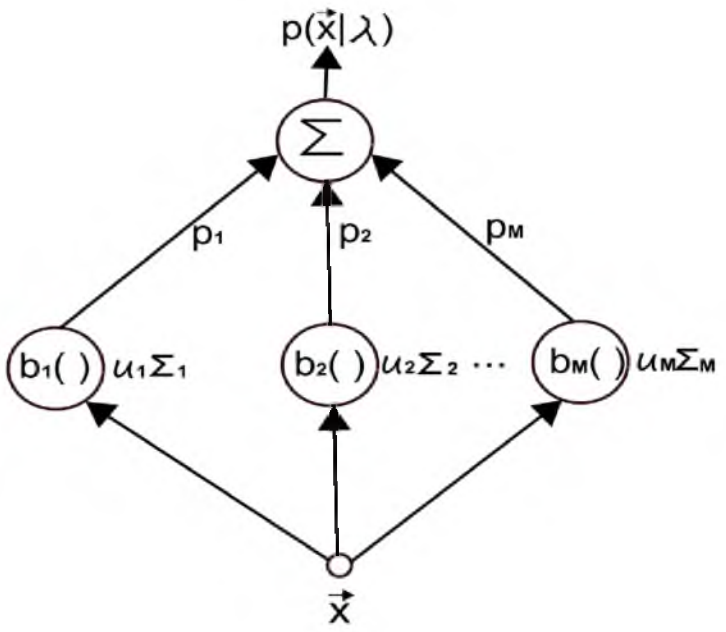

Figura 3.3: Descripción de la densidad de un componente M de una mezcla de gaussianas.

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} \mid \Sigma_i \mid^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)\}$$

Con un vector de medias $\vec{\mu}_i$ y matriz de covarianza Σi. El peso de la mezcla satisface que $\sum_{i=1}^{M} p_i = 1$.

La densidad de toda la mezcla de Gaussianas es parametrizada con el vector de medias, la matriz de covarianzas y la mezcla de medias de todos los componentes de densidad. Estos parámetros se representan a través de:

$$\lambda = \{p_i, \vec{\mu}_i, \Sigma i\} \qquad i = 1, ..., M$$

Para la identificación de locutor, cada locutor es representado por un modelo

GMM y su modelo correspondiente es λ.

- **Algoritmo EM.**

Dado una muestra del habla de un locutor, el objetivo del entrenamiento de su modelo es calcular los parámetros del GMM, λ, los cuales son parecidos a la distribución del vector de características del entrenamiento. Existen muchas técnicas que permiten calcular los parámetros del modelo GMM. El método más común es el cálculo de máxima de verosimilitud, (ML, maximum likelihood).

Lo que busca el método ML es encontrar los parámetros del modelo que maximicen la verosimilitud del modelo, dados los datos de entrenamiento. Para una secuencia de T vectores de entrenamiento $X = \{\vec{x}_i, ..., \vec{x}_T\}$. La verosimilitud del modelo GMM puede ser descrita como:

$$p(X \mid \lambda) = \Pi_{t=1}^{T} p(\vec{x}_t \mid \lambda)$$

Desafortunadamente esta expresión es una función no lineal de los parámetros λ por lo que una maximización directa no puede ser posible. Sin embargo, los parámetros ML calculados pueden obtenerse de manera iterativa usando un algoritmo especial de esperanza máxima (EM, expectation-maximization).

La idea básica del algoritmo EM es que, a partir de un modelo inicial, λ , calcular un nuevo modelo λ , de manera que se cumpla que $p(X \mid \vec{\lambda}) \geq p(X \mid \lambda)$ .El nuevo modelo se convierte así en el modelo inicial para la siguiente iteración, el proceso se repite hasta alcanzar un umbral de convergencia.

En cada una de las iteraciones del algoritmo EM, las siguientes formulas se utilizan para garantizar el incremento del valor de verosimilitud del modelo.

Peso de la mezcla:

$$\vec{p_i} = \frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)$$

Medias:

$$\vec{\mu}_i = \frac{\frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)\vec{x}_t}{\frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)}$$

Varianzas:

$$\vec{\sigma}_i^2 = \frac{\frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)x_t^2}{\frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)} - \vec{\mu}_i^2$$

Donde σ_i^2, x_t, y μ_i, hacen referencia a los elementos de los vectores $\vec{\sigma}_i^2$, $\vec{x}_t$, y $\vec{\mu}_i$, respectivamente.

La razón por la cual se utiliza el modelo GMM se debe a que el trabajo se centrará en la evaluación de las características y no en el modelo. Además, los modelos GMM son más flexibles al clasificar, que un modelo discriminante, en nuestro caso por ejemplo, existen enfermedades que comparten ciertos síntomas; es decir, un síntoma no es necesariamente un síntoma exclusivo de una enfermedad. Si utilizáramos un modelo discrimimante no podríamos modelar los síntomas (o en este caso características de la voz) para todas las enfermedades.

Otra de las razones importantes para seleccionar los modelos generativos es su fácil implementación a través de modelos gráficos como las redes bayesianas. Los modelos gráficos nos permitirán fusionar información y agregar parámetros de diferente índole a los utilizados en esta tesis.

3.1.2. Modelos de Markov Ocultos

Los modelos de Markov Ocultos han sido estudiados desde los 60s por Baum y sus colegas. En los 70s fueron implementados para aplicaciones de procesamiento del habla en el CMU por Baker y en IBM por Jelinek [Rabiner1989]. Su uso dentro del procesamiento del habla no fue la aplicación que se le dio en su inicio, debido a que las primeras publicaciones sobre HMM fueron realizadas en revistas de matemáticas, las cuales no eran leídas por ingenieros. Además de esto, las primeras publicaciones no tenían suficiente teoría de manera que cualquier lector pudiera leerlas, entenderlas y aplicarlas a sus problemas de estudio. Fue Rabiner [Rabiner1989] quien introdujo de manera más amplia los conceptos de los HMM para que estos pudieran ser aplicados al procesamiento del habla. Para explicar cómo funciona un HMM consideremos un sistema descrito en un estado, en cualquier instante de tiempo, de un conjunto de N estados, $S_1, S_2, ..., S^n$, tal como puede observarse en la figura 3.4. (Donde $N = 5$ para simplificar).

En instantes discretos de tiempo, el sistema pasa de un estado a otro (posiblemente se mueva al mismo estado) de acuerdo a un conjunto de probabilidades asociadas a cada uno de los estados. Denotamos los instantes de tiempo asociados a los cambios de estado como $t = 1, 2, ...$, y denotamos el estado actual de tiempo t como q_t. Una descripción de probabilidades más completa del siste-

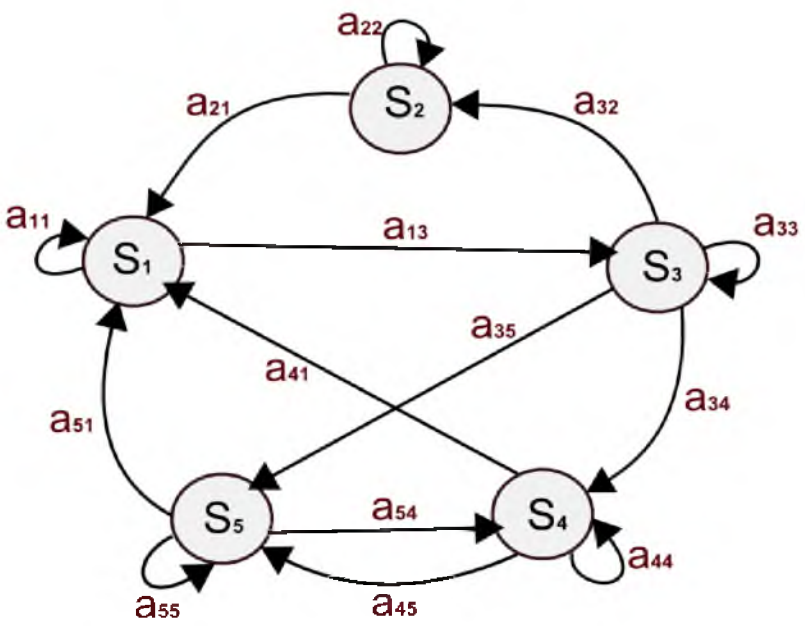

Figura 3.4: Una cadena de Markov de 5 estados (etiquetados desde S_1 a S_2) con estados de transición definidos.

ma, y de manera general requiere una especificación de estado como evaluado o actual (en el tiempo t), así como los predecesores. Para el caso especial de una cadena de Markov discreta y de primer orden, esta descripción de probabilidades es evaluada solo hasta el estado actual y su predecesor, por ejemplo:

$$P[q_t = S_i \mid q_{t-1} = Si, q_{t-2} = S_k, ...] \qquad (1)$$

$$= P[q_t = S_i \mid q_{t-1} = S_i]$$

Además solo consideramos los procesos en los cuales el lado derecho de la ecuación (1) es independiente del tiempo, esto lleva al conjunto de probabilidades

de estados de transición a_{ij} a la forma:

$$a_{ij} = P[q_t = S_i \mid q_{t-1} = S_i], \qquad 1 \leq i, j \leq N$$

Con las siguientes propiedades para los coeficientes de los estados de transición:

$$a_{ij} \geq 0$$

$$\sum a_{ij} = 1$$

Ya que obedecen los estándares estocásticos.

El proceso estocástico anterior puede definirse como un modelo de Markov observable, ya que la salida del proceso es el conjunto de estados en cada instante de tiempo, donde cada estado corresponde a un evento (físico) observable. Otro ejemplo, es el modelo de Markov simple del clima con tres estados. Asumimos que una vez al día (por ejemplo al mediodía), el clima observado es uno de los que a continuación se muestran:

Estado 1: Lluvia o nieve

Estado 2: Nublado

Estado 3: Soleado

Asumimos que el clima en un día t es caracterizado solo por uno de los tres estados mostrados anteriormente, y la matriz A de probabilidades de estados de

transición es:

$$A = \{a_{ij}\} = \begin{pmatrix} 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}2 & 0{,}6 & 0{,}2 \\ 0{,}1 & 0{,}1 & 0{,}8 \end{pmatrix}$$

Dado que el clima en el día 1 (t=1) es soleado (estado 3), podemos hacer la siguiente pregunta: Cual es la probabilidad (de acuerdo al modelo) de que el clima en los siguientes 7 días sea:soleado-soleado-lluvia-lluvia-soleado-nublado-soleado? De manera más formal, definimos la secuencia de observaciones O como $O = \{E_3, E_3, E_1, E_1, E_3, E_2, E_3\}$ que corresponden a $t = 1, 2, 8,$ y deseamos determinar la probabilidad de O, dado el modelo. Esta probabilidad puede expresarse (y evaluarse) como:

$$P(O \mid Modelo) = P[E_3, E_3, E_1, E_1, E_3, E_2, E_3 \mid Modelo]$$

$$= P[E_3]{\cdot}P[E_3 \mid E_3]{\cdot}P[E_3 \mid E_3]{\cdot}P[E_1 \mid E_3]{\cdot}P[E_1 \mid E_1]{\cdot}P[E_3 \mid E_1]{\cdot}P[E_2 \mid E_3]{\cdot}P[E_3 \mid E_2]$$

$$= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23}$$

$$= 1 \cdot (0{,}8)(0{,}8)(0{,}1)(0{,}4)(0{,}3)(0{,}1)(0{,}2)$$

$$= 1{,}536x10^{-4}$$

Donde usamos la notación:

$$\pi_i = P[q_1 = E_i], \qquad 1 \leq i \leq N$$

Hemos considerado modelos de Markov en los cuales cada estado corresponde

a un evento (físico) observable. Este modelo es muy restrictivo para ser aplicado a muchos problemas de interés, por ello se introdujo un concepto más amplio de los modelos de Markov que incluye casos en los que la observación es una función de probabilidad de los estados, denominado Modelo de Markov Oculto. Este modelo es un proceso doblemente estocástico con un proceso estocástico subyacente que no es observable (esta oculto), pero que solo puede ser observado a través de otro conjunto de procesos estocásticos que producen la secuencia de observaciones [Rabiner1989].

Los elementos de un HMM son:

1. N, el número de estados del modelo.
2. M, cantidad de símbolos observables. El símbolo de la observación corresponde a la salida física del sistema que se modela.
3. La distribución de probabilidad de los estados de transición, $A = a_{ij}$, donde

$$a_{ij} = P[q_{t+1} = S_i | q_t = S_i], 1 \leq i, j \leq N$$

4. La distribución de probabilidad de los símbolos observado en el estado $j, B = \{b_j(k)\}$, donde:

$$b_j(k) = P[V_k ent \mid q_t = S_j], 1 \leq j \leq N$$

$$1 \leq k \leq M$$

5. La distribución inicial de estados $\pi = \{\pi_i\}$ donde:

$$\pi_i = P[q_i = S_i], 1 \leq i \leq N$$

Dados los valores apropiados para N, M, A, B y π , el HMM puede ser usado como generador para una secuencia dada de observaciones:

$$O = O_1 O_2 ... O_T$$

Donde cada observación O_T es uno de los símbolos de V y T es el número de observaciones en la secuencia.

3.2. Adaptación de los modelos

La idea básica de la adaptación es obtener el modelo de locutor actualizando los parámetros entrenados en el modelo general, también llamado del mundo, mediante la adaptación. De esta manera se obtiene un acoplamiento más estricto entre el modelo del locutor y el modelo del mundo, además de que permite tener una técnica de cálculo de scoring rápida.

Al igual que el algoritmo EM, la adaptación es un proceso de dos pasos. El primer paso es idéntico al paso de la esperanza del algoritmo EM, donde se calculan las estadísticas suficientes (para el cálculo de los parámetros de la mezcla: peso, media y varianza) de los datos de entrenamiento del locutor y son calculados para cada mezcla del modelo del mundo. A diferencia del segundo paso del algoritmo EM, para la adaptación los nuevos parámetros calculados se combinan

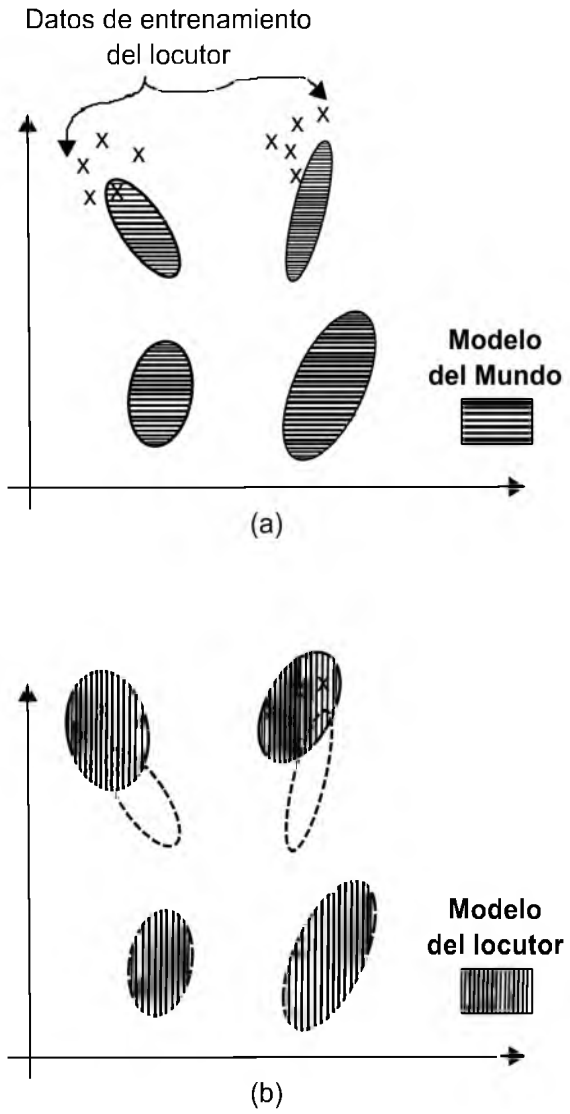

Figura 3.5: Representación gráfica de los dos pasos para la adaptación del modelo de un locutor. (a) El vector de entrenamiento es mapeado probabilisticamente en el modelo del mundo. (b)Los parámetros adaptados de la mezcla se calculan a partir de las estadísticas de los nuevos datos y los parámetros del modelo del mundo. La adaptación depende de los datos, asi que los parámetros del modelo del mundo son adaptados en diferentes proporciones [Reynolds2000].

con los parámetros del modelo del mundo, utilizando un coeficiente de mezcla que depende de los datos. El coeficiente de mezcla se utiliza para que las mez-

clas que tienen pocos datos de entrada del locutor, dependan más de los nuevos parámetros calculados para la estimación final de los parámetros; y las mezclas con pocos datos de entrada del locutor dependan de los antiguos parámetros para la estimación final de parámetros.

La adaptación se realiza de la siguiente manera. Dado un modelo del mundo y un vector de entramiento del locutor, $X = \{x_1, ..., x_T\}$, primero se determinan las probabilidades de alineamiento del vector de entrenamiento con los componentes de la mezcla del modelo del mundo. Esto eso, para cada mezcla i en el modelo del mundo, calculamos:

$$Pr(i \mid x_t) = \frac{\omega_i p_i(x_t)}{\sum_{j=1}^{M} \omega_j p_j(x_t)}$$

Despues utilizamos esta probabilidad $Pr(i \mid x_t)$ y x_t para que a partir de las estadísticas anteriores se calulen los parametros, peso, media y varianza.

$$ni = \sum_{t=1}^{T} Pr(i \mid x_t)$$

$$E_i x = \frac{1}{n_i} \sum_{t=1}^{T} Pr(i \mid x_t) x_t$$

$$E_i(x^2) = \frac{1}{n_i} \sum_{t=1}^{T} Pr(i \mid x_t) x_t^2$$

Esto es igual que el paso de esperanza del algoritmo EM.

Finalmente, estas nuevas estadísticas de los datos de entrenamiento se utilizan para actualizar las estadísticas anteriores del modelo del mundo de la mezcla

i, para realizar la adaptación de los parámetros de la mezcla i con las ecuaciones:

$$\vec{\omega}_i = [\alpha_i^{\omega}\frac{n_i}{T} + (1-\alpha_i^{\omega})\omega_i]\lambda$$

$$\vec{\mu}_i = \alpha_i^{m}E_i(x) + (1-\alpha_i^{m})\mu_i$$

$$\vec{\sigma}_i^2 = \alpha_i^{\nu}E_i(x^2) + (1-\alpha_i^{\nu})(\sigma_i^2 + \mu_i^2)\vec{\mu}_i^2$$

Los coeficientes de adaptación que controlan el balance entre las nuevas y anteriores estimaciones, son $\{\alpha_i^{\omega}, \alpha_i^{m}, \alpha_i^{\nu}\}$ para el peso, las medias y las varianzas respectivamente. El factor de escala, λ es calculado sobre todo el peso de la mezcla adaptada para asegurar que suman la unidad. Notese que las estadísticas, y no los parámetros derivados de ella, como la varianza; son adaptados.

Para cada mezcla y para cada parámetro, un coeficiente de adaptación, que depende de los datos, $\alpha_i^{\rho}, \rho \in \{\omega, m, \nu\}$ se utilizan en las ecuaciones anteriores. Este coeficiente se define como:

$$\alpha_i^{\rho} = \frac{n_i}{n_i + r^{\rho}}$$

donde r^{ρ} (este factor tiene un valor, $r^{\rho} = 16$, calculado experimentalmente) es un factor de mezcla para el parámetro ρ.

Capítulo 4

Protocolo Experimental

4.1. Metodología y Técnicas

La estructura de nuestro sistema para la detección de anomalías en la voz, es muy parecido al sistema utilizado para la identificación de locutor. Los módulos son identicos, solo que el objetivo es diferente. En la identificación de locutor, el objetivo es identificar a una personas a través de las características de su voz, en cambio en la detección de patologías queremos identificar una patología. Debido a estas similitudes, el esquema final del sistema es el mismo. En la figura 4.1. se muestra el esquema del sistema. Además de lo anterior, la adaptación requiere de la identificación del locutor por lo que tener un solo esquema del sistema mejora facilitará la puesta en marcha del sistema final.

A continuación se describen los módulos del sistema y las metodologías desarrolladas en cada una de ellas.

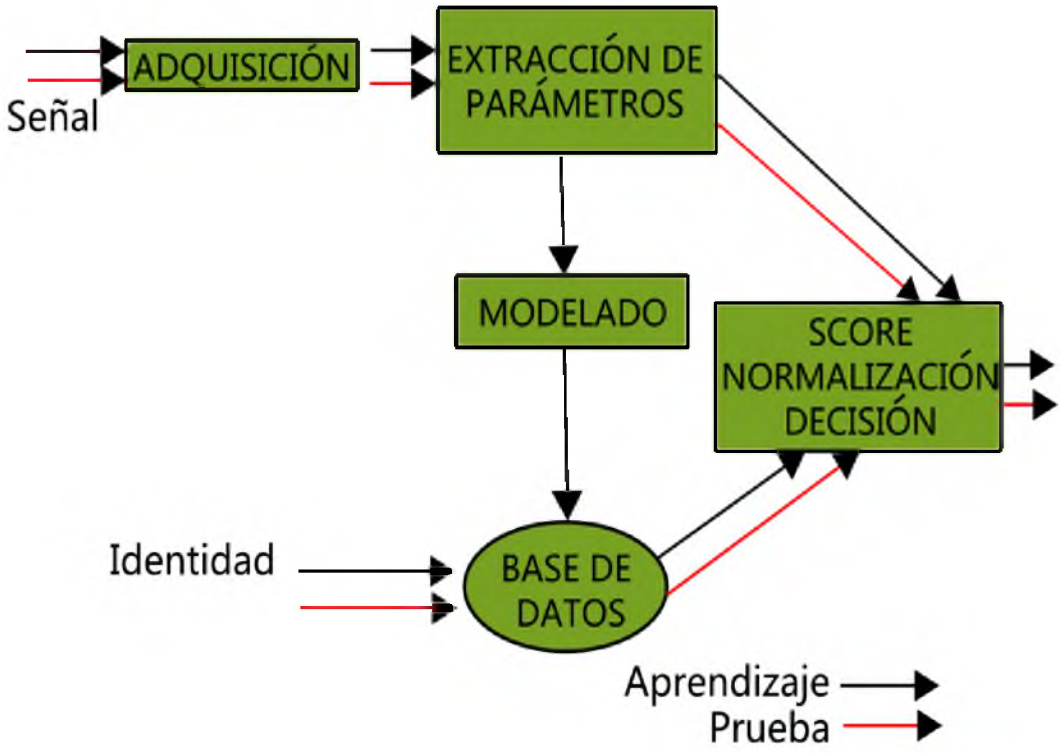

Figura 4.1: Esquema de un sistema para identificacón de locutor.

4.1.1. Adquisición de la voz.

En este módulo, la señal de voz de cada uno de los individuos fue adquirida mediante un micrófono conectado a un convertidor analógico-digital especializado en audio, de manera que el audio fue digitalizado. Este convertidor se conectaba directamente a un puerto USB de una computadora. El archivo de grabación se guardo en formato wav. A la salida de este modulo tenemos la señal de voz digitalizada, y lista para realizar su análisis. En la figura 4.2 se muestra el espectograma de la señal de voz grabada

La base de datos estuvó formada por 4 mujeres y 7 hombres, un total de 11

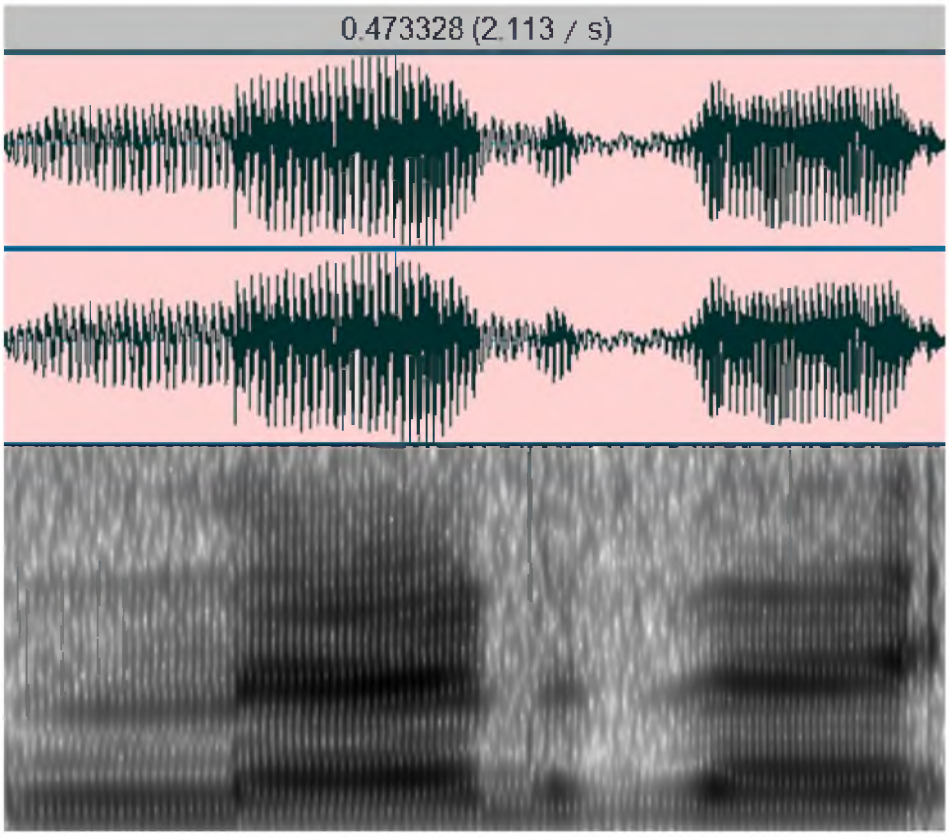

Figura 4.2: Espectograma.

individuos entre 20 y 40 años. Los individuos leyeron un texto de aproximadamente 1000 palabras, dicho texto tenía que ser desconocido por la mayor parte de los individuos. Esta restricción fue principalmente para eliminar el mayor ruido"posible durante la grabación, ya que si los individuos conocían el texto, podían agregarle características de la prosodia, el tono por ejemplo, y las características medidas no serian naturales.

Las muestras de voces se tomaron en dos etapas diferentes, la primera cuando el individuo mostraba alguna patología, y la otra, generalmente dos semanas después; se grabo cuando el individuo ya se encontraba sano.

4.1.2. Extracción de parámetros.

A la entrada de este módulo tenemos la señal de voz obtenida en el módulo de aquisición. Con esta señal se procedió a realizar el análisis correspondiente para determinar las características de interés de la voz (Coeficientes cepstrum, Ritmo, pitch, pendiente del pitch).

Para la extraccíon de los parámetros se utilizaron diferentes herramientas y metodologías, que a continuación se explican. A la salida de este módulo tenemos un archivo que contiene los parámetros calculados, dependiendo cual de ellos haya sido calculado.

4.1.2.1. Coeficientes cepstrum.

Para obtener los coeficientes cepstrum se utiliza la herramienta SPro en la versión 3 [SPRO2009]. Estos coeficientes son la referencia más utilizada en el análisis de la voz debido a que proporcionan información del contenido espectral de la señal [Rabiner1993], por ello son también nuestra referencia.

El esquema de la figura 4.3 muestra como se lleva a cabo el calculo de estos parámetros. Estos parámetros se guardan en un archivo.

4.1.2.2. Pitch.

Otro de los parámetros utilizados fue el pitch, el cual fue calculado utilizando el programa Praat [PRAAT], una herramienta para análisis de señales especializado en audio.

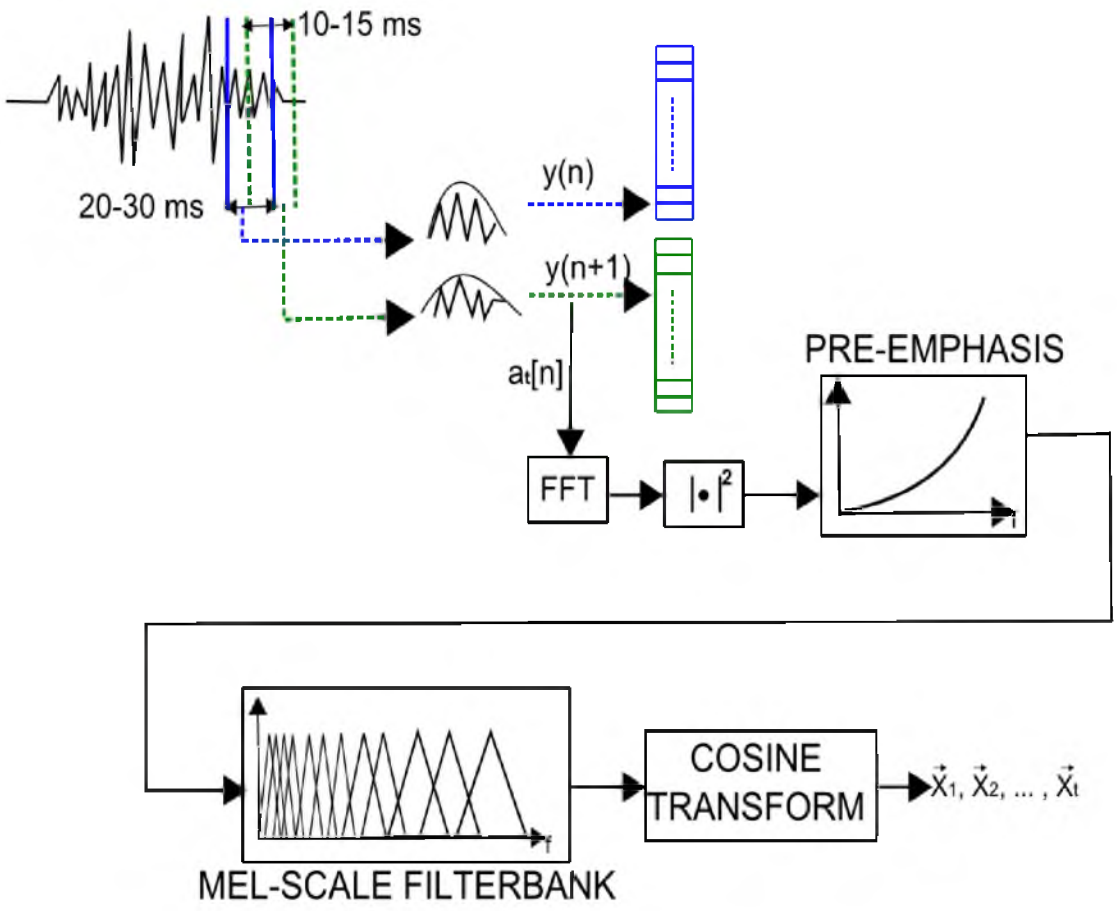

Figura 4.3: Extracción de coeficientes cepstrales.

El pitch es la frecuencia que determina el tono de voz de cada persona, es decir, es la frecuencia que nos caracteriza y permite que nos identifiquen al hablar, deriva directamente de las cuerdas vocales. La principal razón de utilización del pitch en nuestro proyecto es que, este parámetro conlleva información de la glotis (abertura entre las cuerdas vocales), y como hemos dicho anteriormente, los individuos enfermos presentan gripa, con lo que se espera se tengan buenos resultados.

Dentro del programa Praat, existen varios métodos para el cálculo del pitch, el método que se utilizó fue el de cross-correlación para minimizar el tiempo de cálculo puesto que nuestro sistema debe trabajar en tiempo real, y los otros métodos requieren más tiempo para el cálculo del pitch. Existen métodos más rápidos en tiempo de cálculo, sin embargo los valores obtenidos no son lo suficientemente buenos para nuestro trabajo. El método de cross-correlación es un buen compromiso entre la calidad de los valores y el tiempo de computo.

Estos parámetros se almacenan, de la misma forma que los coeficientes cepstrum, en un archivo.

4.1.2.3. Pendiente del Pitch.

Durante la experimentación se observó que el pitch de las personas enfermas es diferente al de las personas sanas, tiene ciertas variaciones. En las figuras 4.4 y 4.5 se muestran estas variaciones.

En las personas enfermas, el pitch es mas plano, la gráfica se asemeja más a una línea recta, y en los individuos sanos presenta curvas más pronunciadas. Por lo que se propusó como un nuevo parámetro de clasificación.

Para medir las variaciones en el pitch, se calcularon sus pendientes. Estas pendientes se obtuvieron utilizando la siguiente ecuación:

$$y_i^{'}(t) = \frac{2y_i(t+2) + y_i(t+1) - y_i(t-1) - 2y_i(t-2)}{10}$$

Lo que corresponde a la primera derivada del picth.

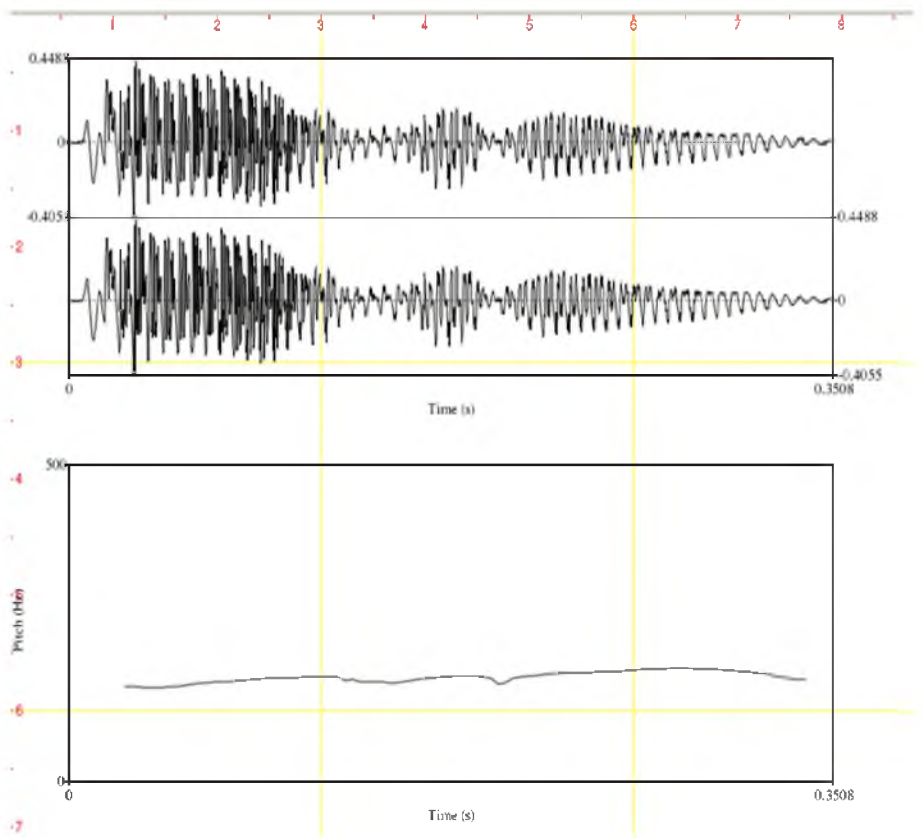

Figura 4.4: Pitch de una persona sana.

4.1.3. Modelado.

Los archivos generados en el modulo de extracción de parámetros son la entrada de este módulo, a partir de ellos se procedió con la creación del modelo del sistema.

En este módulo se crearon los modelos GMM descritos anteriormente. Los modelos fueron generados a través del programa BECARS [BECARS], una herramienta especializada en audio. El programa utiliza el algoritmo EM para calcular los parámetros (media, varianza) de las Gaussianas dentro de la mezcla.

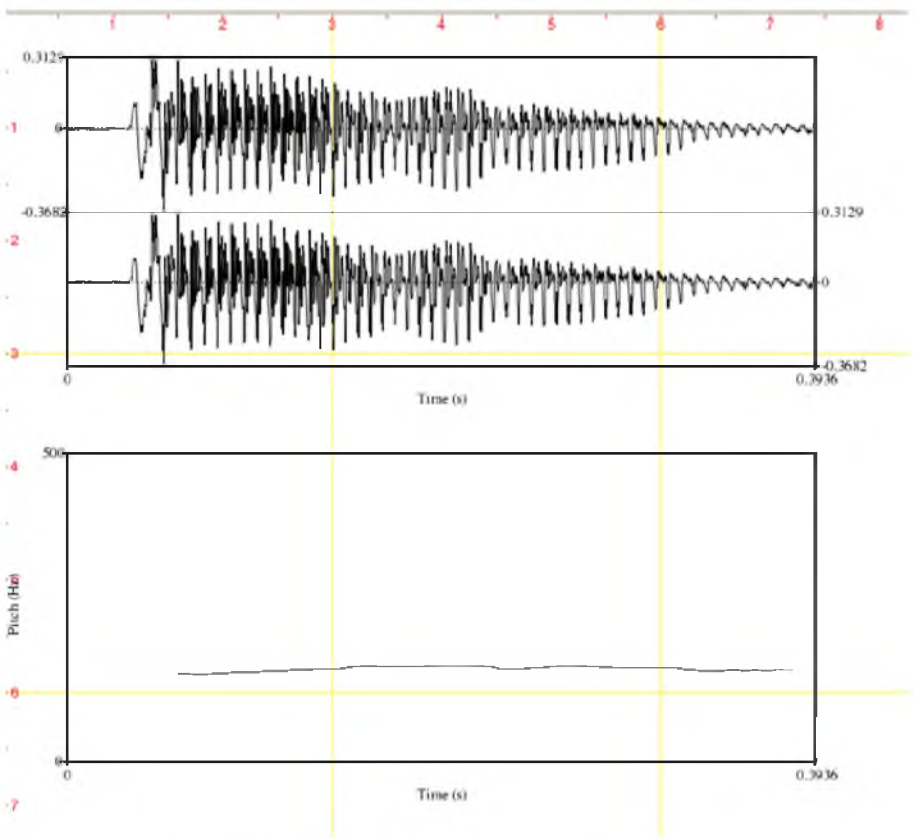

Figura 4.5: Pitch de una persona enferma.

Con dicho programa también se realizó la adaptación del modelo.

La creación de los modelos se realizó en dos etapas, la primera sin la adaptación del modelo de los individuos, y la segunda, utilizando adaptación, en las dos etapas se utilizó el mismo programa.

Para los modelos sin adaptación se crearon tres modelos del mundo para cada uno de los parámetros (cepstrum, pitch, pendiente del pitch), uno para individuos sanos, uno para los individuos enfermos y uno más donde se encontraban todos

los individuos (enfermos y sanos). En el caso del pitch se hizo una separación por géneros, ya que la frecuencia difiere según sea el caso, es decir, la frecuencia en las mujeres es más alta que la frecuencia de los hombres como ya se explicó en el capítulo de producción de la voz.

En los modelos adaptados se crearon cinco modelos del mundo para todos los parámetros, ya que en estos se separaron los individuos por géneros, de tal forma que se obtuvieron dos modelos para los individuos sanos y dos para los individuos enfermos, más el modelo del mundo con todos los individuos. Los modelos con los individuos separados se utilizaron para adaptar el modelo de cada uno de los individuos que conforman la base de datos.

Debido a esto, durante esta etapa se hizo necesario realizar una identificación de locutor para saber de quién es la voz que se está adaptando.

Una vez creados los modelos, se compararon los parámetros de entrada con el modelo correspondiente (enfermos o sanos) y calcular su score, un valor que nos dice que tanto se parecen los parámetros al modelo. Con estos scores se clasificaron las voces como sanas o patológicas. Por ejemplo, tenemos como entrada del sistema, los parámetros de la voz de un individuo X, estos parámetros se comparan tanto con el modelo de individuos sanos, como el modelo de individuos enfermos, con lo que obtenemos dos scores,y dependiendo de cual score es el mayor es como se clasifica la voz.

4.1.4. Reconocimiento de Locutor.

Es necesario realizar un reconocimiento de locutor durante el análisis, debido a la adaptación individualizada de los modelos. Para el reconocimiento de locutor se utilizó la plataforma Becars.

Capítulo 5

Evaluación y Resultados

En este capítulo se analizan los porcentajes de clasificación que tuvieron los parámetros utilizados para la clasificación de la voz en normal o patológica.

5.1. Scores obtenidos con los diferentes parámetros.

5.1.1. Coeficientes cepstrum.

En los coeficientes cepstrum se utilizaron 64 gausianas para la creación de los modelos del mundo para los individuos enfermos y los individuos sanos. Los scores obtenidos se muestran en las tablas 5.1 y 5.2 [Velázquez2011O].

Donde los números en rojo, marcan las clasificaciones incorrectas y los números en negro representan las clasificaciones correctas.

Cuadro 5.1: TCS Scores para individuos sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,008547526	0,987441323
Individuo 2	1,009668465	0,986197037
Individuo 3	0,977494953	1,022245822
Individuo 4	1,019558003	0,969836977
Individuo 5	1,012775057	0,991999511
Individuo 6	0,976497236	1,022049263
Individuo 7	0,975178402	1,019990631
Individuo 8	1,0160064	0,975176499
Individuo 9	0,978298426	1,021843205
Individuo 10	1,00411677	1,000399573
Individuo 11	0,989609467	1,01584361

Cuadro 5.2: TCE Scores para individuos enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,004412804	1,005016205
Individuo 2	1,003157699	0,993768985
Individuo 3	1,015478949	0,994456748
Individuo 4	1,029912825	0,979153259
Individuo 5	1,00715611	0,992658941
Individuo 6	0,971983239	1,025024008
Individuo 7	0,979050101	1,017874466
Individuo 8	1,025282991	0,974413698
Individuo 9	0,966587362	1,028471152
Individuo 10	0,984295674	1,01440983
Individuo 11	1,037065675	0,970733003

En la clasificación de individuos sanos, se obtuvo un 45 % de clasificación correcta y un 55 % de clasificación errónea, como puede observarse en la figura 5.1. Los porcentajes de clasificación en el caso de los individuos enfermos se invir-

tieron comparados con los porcentajes de los individuos sanos, es decir, se tuvo una clasificación correcta del 55 % y una clasificación errónea que correspondió a un 45 %.

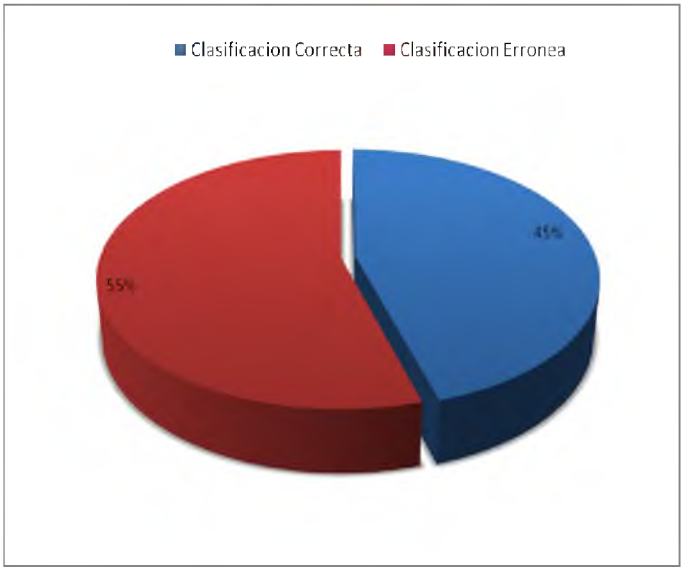

Figura 5.1: Porcentajes de clasificación utilizando coeficientes cepstrum para individuos sanos (Sin Adapt.).

5.1.2. Pitch.

Para el pitch se utilizaron 2 gausianas para la creación de los modelos del mundo en los individuos enfermos y los individuos sanos.

Para este parámetro en específico se crearon dos modelos (enfermos y sanos). Esto se debe a que se separaron las mujeres y los hombres dadas las frecuencias naturales del pitch (frecuencias de las mujeres es más alto que las frecuencias del sonido de las voces de los hombres).

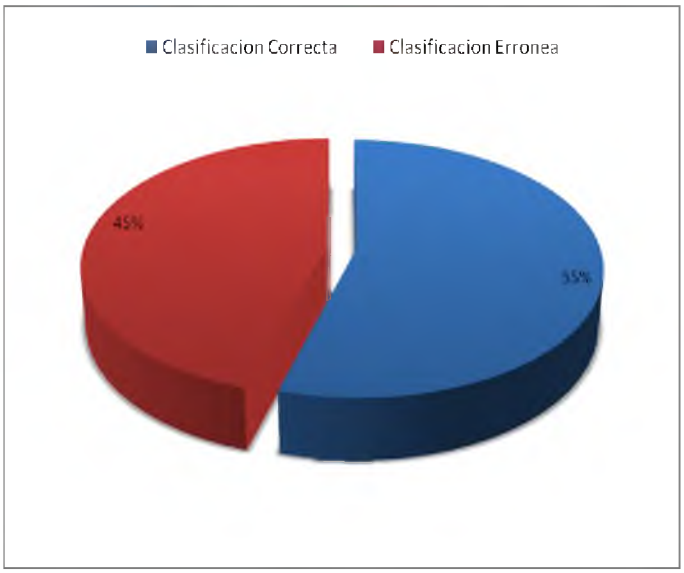

Figura 5.2: Porcentajes de clasificación utilizando coeficientes cepstrum para individuos enfermos (Sin Adapt.).

Los scores obtenidos se muestran en las tablas 5.3 y 5.4 [Velázquez2011H]. Al calcular los scores, utilizando el pitch, se obtuvo un individuo, dentro del grupo de los enfermos, cuyo score es el mismo al compararlo tanto con el modelo de los enfermos y el modelo de los sanos.

De manera gráfica podemos observar los resultados de las clasificaciones en las figuras 5.3 y 5.4. En ellas podemos observar que el pitch tiene una mejor clasificación en los individuos sanos que en los individuos enfermos. Esto es, al clasificar los individuos, se tuvo un menor porcentaje de clasificación erróneo para los individuos sanos y en los individuos enfermos un porcentaje mayor.

5.1.3. Pendiente del pitch.

Para la pendiente del pitch se utilizaron 2 gausianas (el mismo número que para el picth) para la creación de los modelos del mundo en los individuos enfer-

Cuadro 5.3: TPS Scores para individuos sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	0,999980	0,999977
Individuo 2	0,999567	0,999168
Individuo 3	1,000041	1,000054
Individuo 4	1,000089	1,000064
Individuo 5	1,000057	1,000037
Individuo 6	1,000124	1,000215
Individuo 7	1,000167	1,000218
Individuo 8	0,999955	1,000009
Individuo 9	0,999941	1,000001
Individuo 10	1,000122	1,000183
Individuo 11	0,999969	1,000020

Cuadro 5.4: TPE Scores para individuos enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	0,999980	0,999987
Individuo 2	0,999591	0,999272
Individuo 3	0,999980	1,000018
Individuo 4	1,000405	1,000228
Individuo 5	0,999980	0,999991
Individuo 6	1,000086	1,000177
Individuo 7	1,000176	1,000212
Individuo 8	1,000090	1,000090
Individuo 9	1,000086	1,000074
Individuo 10	1,000070	1,000043
Individuo 11	0,999959	1,000026

mos y los individuos sanos.

Para este parámetro el número de gausianas es igual a 2 debido a que las varia-

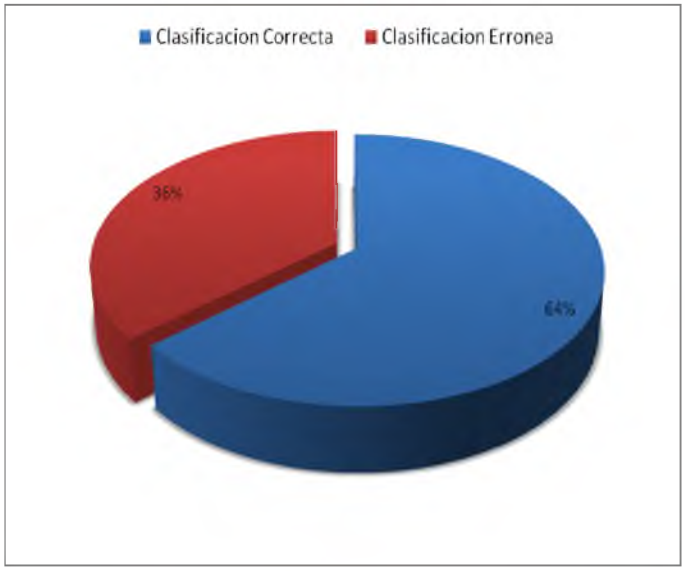

Figura 5.3: Porcentajes de clasificación utilizando el pitch para individuos sanos (Sin Adapt.).

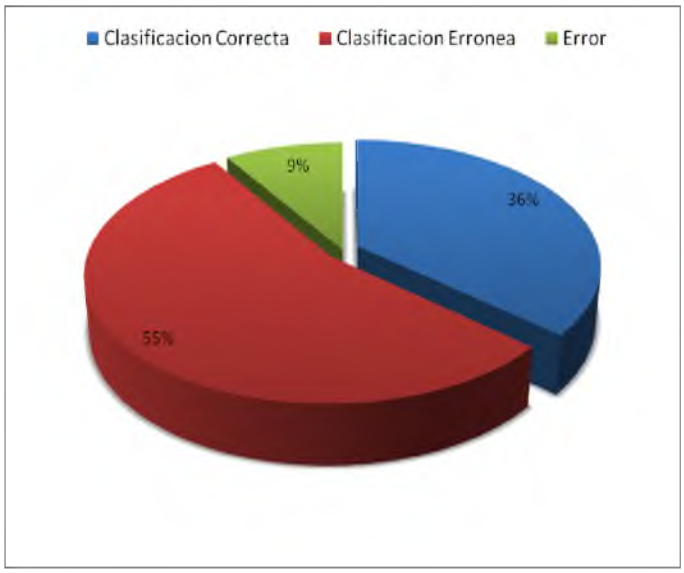

Figura 5.4: Porcentajes de clasificación utilizando el pitch para individuos enfermos (Sin Adapt.).

ciones que tiene la pendiente del picth es mínima.

Los scores obtenidos se muestran en las tablas 5.5 y 5.6.

Cuadro 5.5: TPPS Scores para individuos sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000004	0,999996
Individuo 2	1,000001	0,999996
Individuo 3	1,000003	0,999996
Individuo 4	1,000000	0,999998
Individuo 5	1,000002	0,999997
Individuo 6	0,999998	1,000001
Individuo 7	0,999999	0,999999
Individuo 8	1,000004	0,999995
Individuo 9	1,000002	0,999996
Individuo 10	1,000001	0,999999
Individuo 11	1,000004	0,999996

Cuadro 5.6: TPPE Scores para individuos Enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000003	0,999996
Individuo 2	1,000005	0,999994
Individuo 3	1,000004	0,999996
Individuo 4	1,000001	0,999998
Individuo 5	1,000004	0,999996
Individuo 6	1,000000	0,999999
Individuo 7	0,999999	1,000000
Individuo 8	1,000002	0,999996
Individuo 9	1,000003	0,999996
Individuo 10	1,000001	0,999997
Individuo 11	1,000002	0,999996

De la misma forma que sucedió cuando se utilizo el pitch, al utilizar la pendiente como parámetro se obtuvo un individuo cuyo score es el mismo al compararlo

con el modelo de los enfermos y con el modelo de los sanos, sin embargo, esta vez el individuo se encuentra dentro del grupo de los sanos.

Los porcentajes de clasificación de acuerdo con las tablas anteriores pueden observarse en la figuras 5.5 y 5.6. En estas figuras podemos observar que la pendiente del pitch clasificó mejor a los individuos enfermos que a los individuos sanos. Es decir, el porcentaje de clasificación erróneo fue menor cuando se clasificaron los individuos enfermos y mayor cuando se clasificaron los individuos sanos.

Figura 5.5: Porcentajes de clasificación utilizando la pendiente del pitch para individuos sanos (Sin Adapt.).

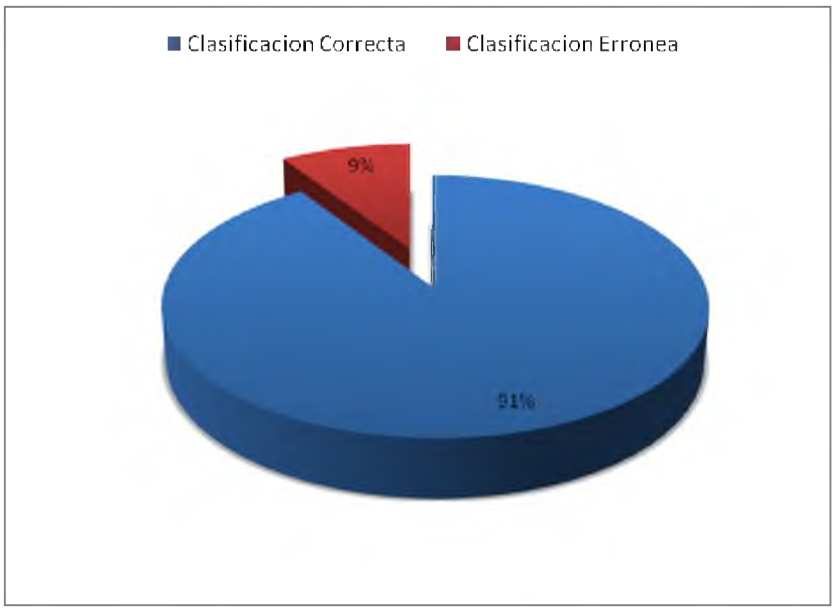

Figura 5.6: Porcentajes de clasificación utilizando la pendiente del pitch para individuos enfermos (Sin Adapt.).

5.1.4. Comparación de Resultados.

Para los individuos sanos, el parámetro que tuvo un mayor porcentaje correcto de clasificación es la pendiente del pitch, el cual corresponde a un 81 % de clasificación correcta, tal como puede observarse en la figura 5.7. La razón de estos resultados se debe principalmente a que el pitch lleva información relacionada con la glotis.

En los individuos enfermos el parámetro que tuvo un mayor porcentaje de clasificación correcta fue la pendiente del pitch, el cual tuvo un porcentaje de 91 %. Podemos observar esto de manera grafica en la figura 5.8 que a continuación se muestra.

De igual manera, de los tres parámetros utilizados podemos observar, que los

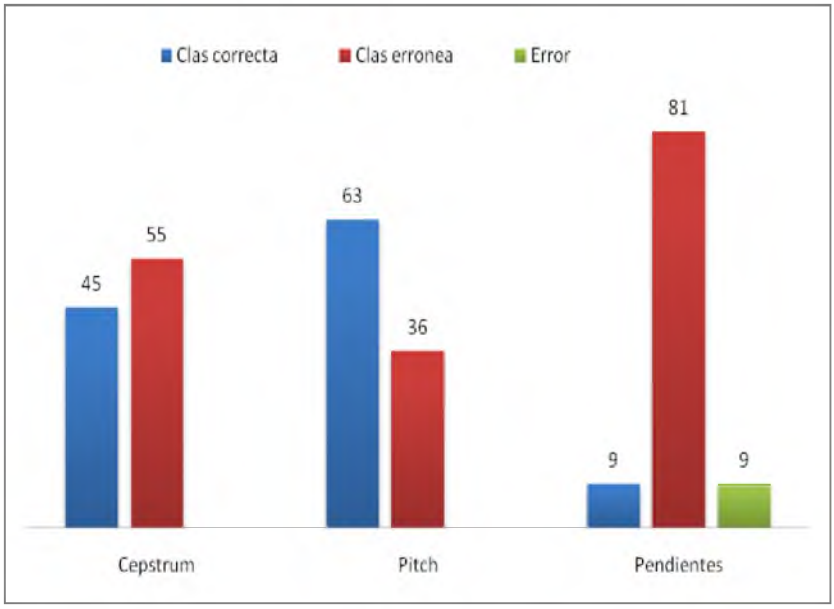

Figura 5.7: Porcentajes de clasificación de los parámetros para individuos sanos (Sin Adapt.).

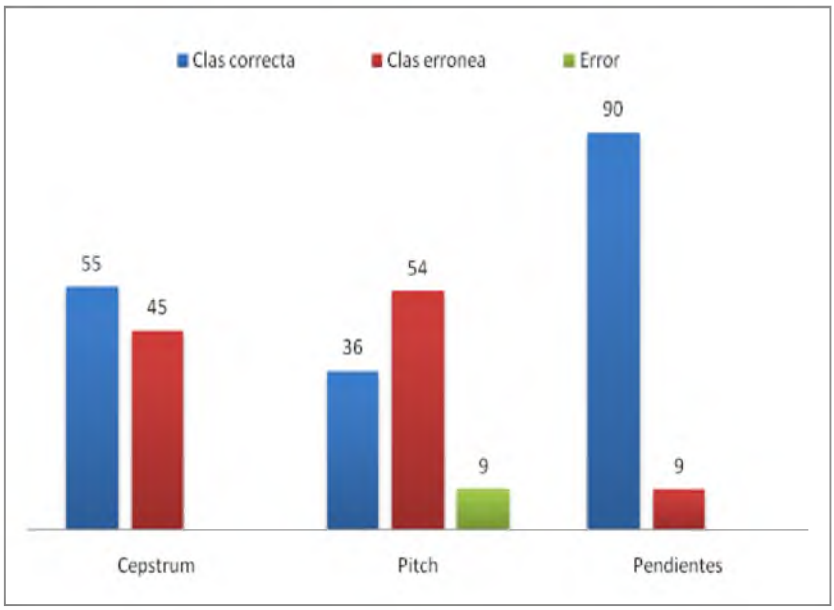

Figura 5.8: Porcentajes de clasificación de los parámetros para individuos enfermos (Sin Adapt.).

coeficientes cepstrum son los que tuvieron un nivel de desempeño más bajo. Esto se debe a que a diferencia del pitch, los coeficientes cepstrum llevan información del conducto vocal, como la enfermedad tratada es gripe por eso se obtienen mejores clasificaciones al utilizar el pitch o la pendiente del mismo que cuando se utilizan los coeficientes cepstrum.

Se pudo observar que tanto en el pitch como en las pendientes de este, hubo individuos que presentaron scores iguales al compararlos con los modelos de enfermos y sanos, esto es, no se pudieron clasificar como enfermos o como sanos debido a que su valor de verosimilitud fue el mismo en ambos casos.

5.2. Scores obtenidos con los modelos adaptados.

Los modelos de cada uno de los individuos fueron adaptados a partir de los modelos del mundo de los individuos sanos y los individuos enfermos para obtener mejores resultados, ya que los porcentajes de clasificación obtenidos con los modelos sin adaptar no son tan satisfactorios como se esperaba.

A continuación se muestran los resultados de los scores obtenidos al adaptar los modelos utilizados..

5.2.1. Coeficientes cepstrum.

En los coeficientes cepstrum, se utilizaron 64 gausianas para la creación de los modelos del mundo para los individuos enfermos y los individuos sanos, tanto

para hombre como para mujeres. Los scores obtenidos se muestran en las tablas 5.7 y 5.8.

Donde los números en rojo, marcan las clasificaciones incorrectas y los números en negro representan las clasificaciones correctas.

Cuadro 5.7: TCSA Scores para individuos Sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,008440334	0,988840335
Individuo 2	1,021336592	1,020398813
Individuo 3	1,009334232	1,052513231
Individuo 4	1,030670546	1,117982899
Individuo 5	1,006394826	0,98195845
Individuo 6	1,021412836	1,091380664
Individuo 7	1,014766996	1,083477758
Individuo 8	1,063169451	0,959443357
Individuo 9	1,033491975	1,071913614
Individuo 10	1,016422533	1,012049236
Individuo 11	0,990579278	1,004954313

En la clasificación de individuos sanos, se obtuvo un 55 % de clasificación correcta y un 45 % de clasificación errónea, como puede observarse en la figura 5.9.

Los porcentajes de clasificación en el caso de los individuos enfermos fueron los mismos que los porcentajes obtenidos al clasificar a los individuos sanos.

Comparados con los resultados de los scores obtenidos utilizando modelos sin adaptar, estos resultados son mejores, en un 10 %. En el caso anterior, utilizando los modelos sin adaptar se tenía un mayor porcentaje de error.

Cuadro 5.8: TCEA Scores para individuos Enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,012729434	1,012950542
Individuo 2	1,026132563	1,000070391
Individuo 3	1,018454391	1,008508277
Individuo 4	1,015485769	0,950881604
Individuo 5	1,025398627	0,997312871
Individuo 6	1,0327874	1,097569893
Individuo 7	1,013688664	1,071971792
Individuo 8	1,046943498	0,951854453
Individuo 9	1,033152343	1,055637117
Individuo 10	1,020587707	1,061433502
Individuo 11	1,046835961	0,902498191

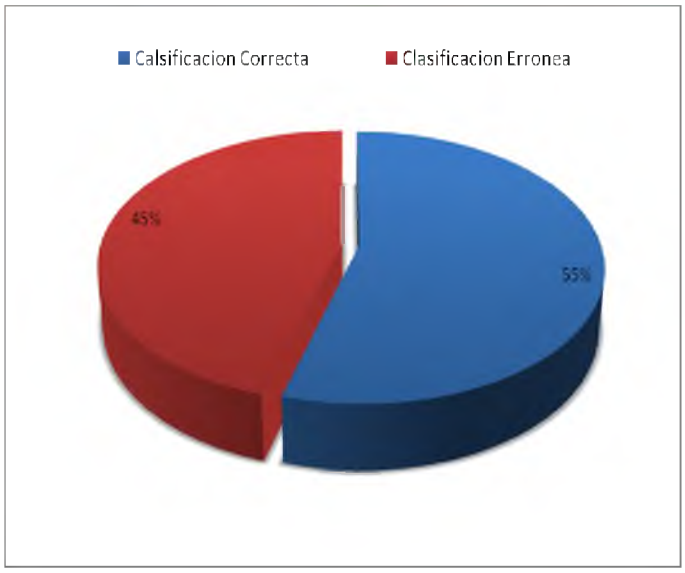

Figura 5.9: Porcentajes de clasificación utilizando coeficientes cepstrum para individuos sanos (Adaptados).

Esta mejora es resultado de una mejor modelación de los parámetros, puesto que, por ejemplo, para crear el modelo de un individuo enfermo, utilizamos como

referencia el modelo del mundo de los individuos enfermos, de manera que el modelo resultante se parezca más al modelo general de individuos enfermos.

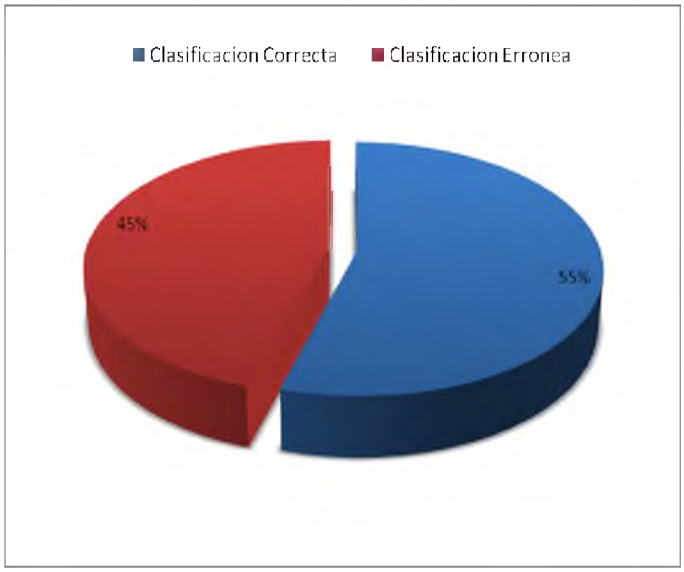

Figura 5.10: Porcentajes de clasificación utilizando coeficientes cepstrum para individuos enfermos (Adaptados).

5.2.2. Pitch.

Para el pitch se utilizaron 2 gausianas para la creación de los modelos del mundo en los individuos enfermos y los individuos sanos.

Los scores obtenidos se muestran en las tablas 5.9 y 5.10.

Al calcular los scores, utilizando el pitch, se obtuvo un individuo, dentro del grupo de los sanos, cuyo score es el mismo al compararlo tanto con el modelo de los enfermos y el modelo de los sanos.

De manera gráfica podemos observar los resultados de las clasificaciones en las figuras 5.11 y 5.12. En ellas podemos observar que el pitch tiene porcen-

Cuadro 5.9: TPSA Scores para individuos Sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000010	1,000028
Individuo 2	1,002904	0,099751
Individuo 3	1,000015	1,000015
Individuo 4	0,999903	0,999996
Individuo 5	0,999990	0,999993
Individuo 6	1,000164	1,000223
Individuo 7	1,000192	1,000201
Individuo 8	0,999976	1,000063
Individuo 9	1,000012	1,000074
Individuo 10	1,000033	1,000152
Individuo 11	1,000053	1,000063

Cuadro 5.10: TPEA Scores para individuos Enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000015	1,000008
Individuo 2	1,002466	1,002549
Individuo 3	1,000036	1,000034
Individuo 4	1,000224	0,999939
Individuo 5	1,000003	1,000009
Individuo 6	1,000118	1,000173
Individuo 7	1,000177	1,000160
Individuo 8	1,000037	0,999981
Individuo 9	1,000002	0,999964
Individuo 10	1,000004	0,999930
Individuo 11	1,000077	1,000076

tajes de clasificación aceptables, a pesar de presentar un error de clasificación en los individuos sanos, tiene mejores resultados que los coeficientes cepstrum. Además de que en los individuos enfermos no se presentó ningún error de clasi-

ficación como en el caso de los modelos sin adaptar.

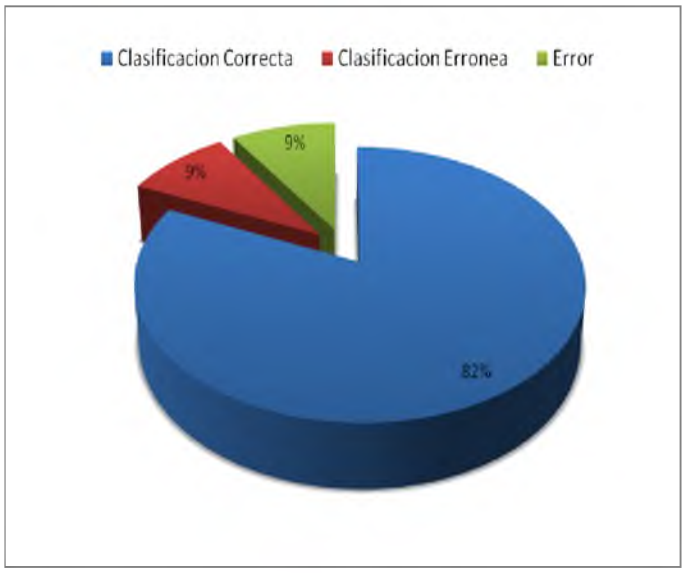

Figura 5.11: Porcentajes de clasificación utilizando el pitch para individuos sanos (Adaptados).

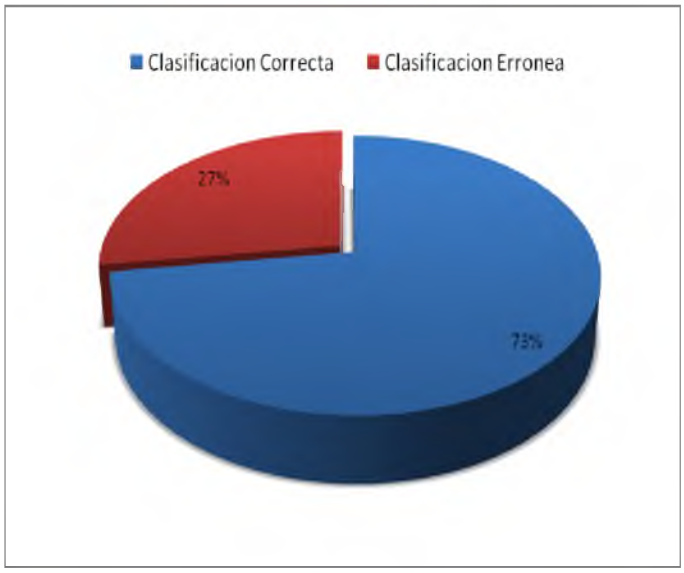

Figura 5.12: Porcentajes de clasificación utilizando el pitch para individuos Enfermos (Adaptados).

5.2.3. Pendiente del pitch.

Para la pendiente del pitch se utilizaron 2 gausianas (el mismo número que para el picth) para la creación de los modelos del mundo en los individuos enfermos y los individuos sanos.

Para este parámetro el número de gausianas es igual a 2 debido a que las variaciones que tiene la pendiente del picth son mínimas.

Los scores obtenidos se muestran en las tablas 5.11 y 5.12.

Cuadro 5.11: TPPSA Scores para individuos Sanos.

Individuos Sanos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000046	1,000087
Individuo 2	0,999975	0,999997
Individuo 3	0,999998	0,999996
Individuo 4	1,000126	1,000121
Individuo 5	1,000005	1,000012
Individuo 6	1,000008	1,000016
Individuo 7	1,000038	1,000037
Individuo 8	0,999999	0,999994
Individuo 9	1,000049	1,000056
Individuo 10	1,000000	1,000005
Individuo 11	0,999999	1,000006

Los porcentajes de clasificación de acuerdo con las tablas anteriores pueden observarse en las figuras 5.13 y 5.14. En estas figuras podemos observar que la pendiente del pitch clasificó mejor a los individuos enfermos que a los individuos sanos.

Cuadro 5.12: TPPEA Scores para individuos Enfermos.

Individuos Enfermos		
	GMMEnfermos	**GMMSanos**
Individuo 1	1,000015	1,000009
Individuo 2	1,000039	0,999980
Individuo 3	1,000034	1,000001
Individuo 4	1,000126	1,000123
Individuo 5	1,000008	0,999996
Individuo 6	1,000009	1,000007
Individuo 7	1,000067	1,000057
Individuo 8	1,000008	0,999998
Individuo 9	1,000032	1,000029
Individuo 10	1,000000	0,999999
Individuo 11	1,000002	0,999989

En la clasificación de los individuos enfermos no se tuvo ningún error, es decir, su porcentaje de clasificación fue del 100 %. En cambio, para la clasificación de los individuos sanos, se obtuvo un porcentaje de error del 36 %.

5.2.4. Comparación de Resultados.

Para los individuos sanos, el parámetro que tuvo un mayor porcentaje correcto de clasificación fue el pitch, el cual corresponde a un 82 % de clasificación correcta, tal como puede observarse en la figura 5.15.

Para los individuos enfermos el parámetro que tuvo un mayor porcentaje de clasificación correcta fue la pendiente del pitch, el cual tuvo un porcentaje del 100 %, es decir que no tuvo errores al clasificar a los individuos como enfermos. Podemos observar esto de manera grafica en la figura 5.16.

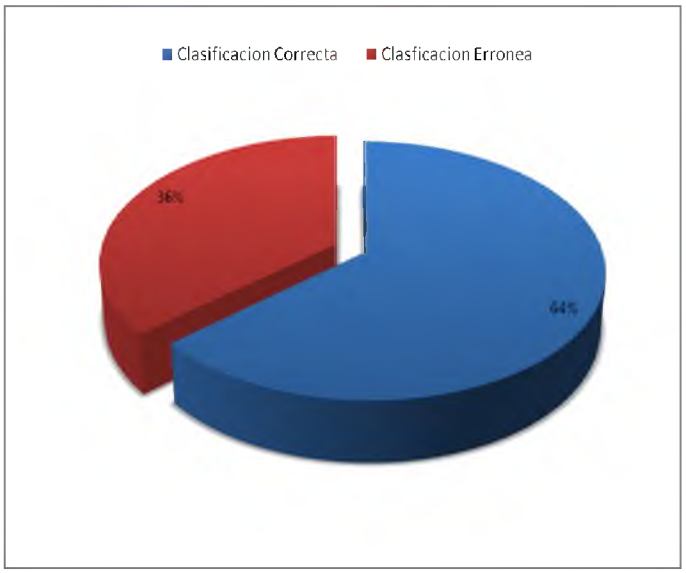

Figura 5.13: Porcentajes de clasificación utilizando la pendiente del pitch para individuos sanos (Adaptados).

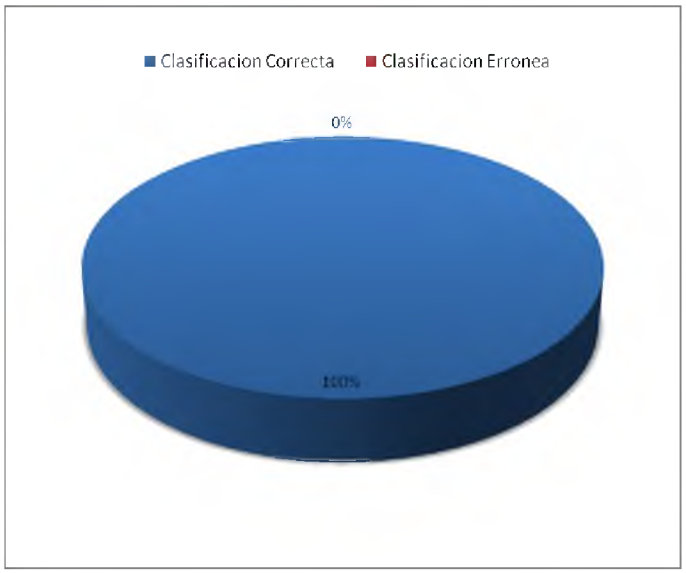

Figura 5.14: Porcentajes de clasificación utilizando la pendiente del pitch para individuos enfermos (Adaptados).

De los tres parámetros, el pitch y la pendiente son los parámetros que mejor

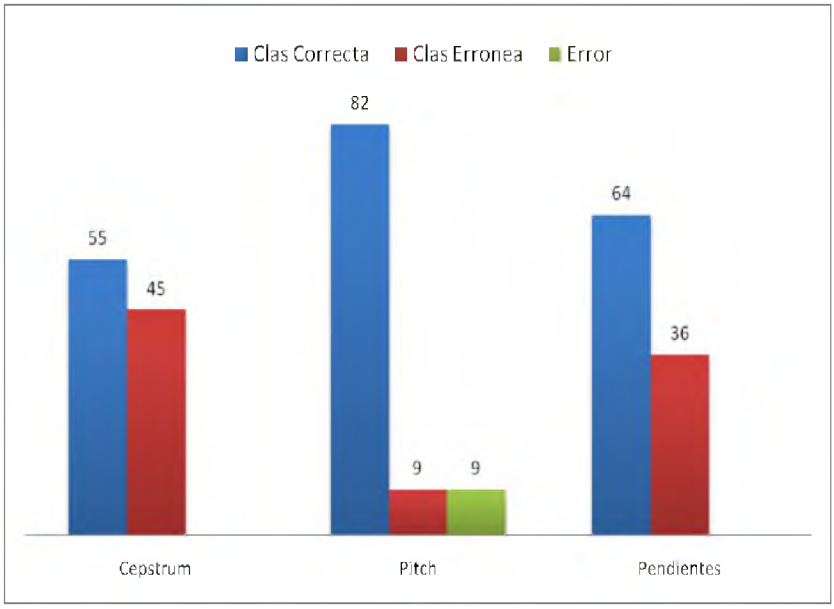

Figura 5.15: Porcentajes de clasificación de los parámetros para individuos sanos (Adaptados).

desempeño tuvieron al clasificar a los individuos.

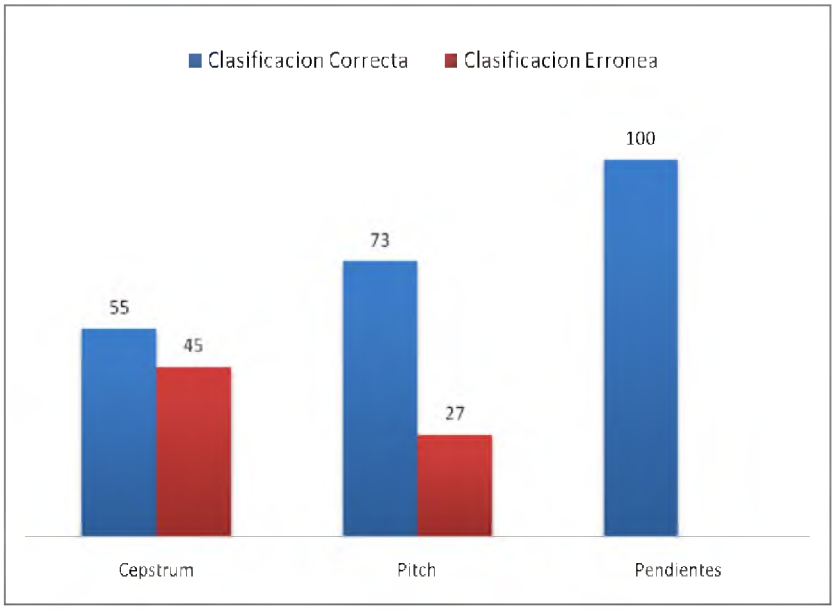

Figura 5.16: Porcentajes de clasificación de los parámetros para individuos enfermos (Adaptados).

De igual manera, de los tres parámetros utilizados podemos observar, que los coeficientes cepstrum son los que tuvieron un nivel de desempeño más bajo. A pesar de que los porcentajes de clasificación erróneos fueron menores a los porcentajes de clasificación correcta, no hay mucha diferencia entre ellos.

Los resultados sobre el desempeño de los coeficientes cepstrum comparado con el pitch y la pendiente del pitch, son debidos a que la diferencia en la información que llevan cada uno de ellos es muy significativa.

Los resultados obtenidos utilizando la adaptación de los modelos en general, es mejor, ya que los tres parámetros tuvieron mejorías. Y dados estos resultados, se espera que con la combinación del pitch y las pendientes se obtengan aun mejores resultados; pues fueron los parámetros que obtuvieron los mejores resultados, y además, los resultados variaron entre individuos sanos y enfermos, esto es, uno tuvo mejores resultados al clasificar individuos enfermos y otro al clasificar individuos sanos.

Capítulo 6

Conclusiones y perspectivas

En el presente trabajo se realizó el análisis acústico de la voz para determinar las características en ella, que nos permitan identificar alguna enfermedad (aquellas relacionadas directamente con el aparato fonatorio).

Para probar los parámetros seleccionados se utilizó el modelo de mezcla de Gaussianas. Se seleccionó este modelo debido a que, se buscaba probar el desempeño de los parámetros y no el modelo en sí. Por otra parte, un GMM es fácil de implementar, además de permitir la adaptación conforme se van agregando nuevos datos.

Se utilizaron tres parámetros durante los experimentos, los coeficientes cepstrum, el pitch y la pendiente del pitch. Se compararon los resultados obtenidos al utilizar cada uno de ellos, y los que obtuvieron mejores resultados fueron el pitch y la pendiente del pitch. Más adelante, se realizó la adaptación de los modelos con lo cual los resultados mejoraron significativamente.

A raíz de estos experimentos, podemos concluir que los parámetros que funcionan mejor al clasificar la señal de voz son el pitch y la pendiente del pitch; ya que ellos obtuvieron los mejores resultados, uno para clasificar a los individuos sanos y el otro para clasificar a los individuos enfermos. Esto se puede explicar, debido a que el picth conlleva información relacionada con la glotis, y como, la enfermedad utilizada es una enfermedad respiratoria por eso son mejores.

Los resultados obtenidos al utilizar la pendiente del pitch fueron mejores respecto a los obtenidos con cualquiera de los otros parámetros.

Al proponer un nuevo parámetro para el modelado de la voz, la pendiente del pitch, la cual presentó resultados satisfactorios durante los experimentos realizados, hemos cumplido con los objetivos que nos marcamos inicialmente.

Al mismo tiempo, se aportan las bases para la creación de un sistema completo de prevención en materia de salud al incluir en nuestro estudio un texto de aproximadamente 1000 palabras, de manera que se obtuviera una muestra más significativa.Esto es, el sistema propuesto es independiente completamente del texto, lo que lo hace más general.

Como líneas de trabajo futuro, se realizará la fusión de información (pitch y la pendiente del pitch) mediante modelos gráficos para mejorar los resultados.
El siguiente paso es crear un sistema de reconocimiento de voz para agregar información suprasegmental de alto nivel y poder identificar otro tipo de patologías.

Bibliografía

[Álvarez2001] Agustín Álvarez Marquina. ***Fundamentos del Reconocimiento Automático de la voz: Historia de los Sistemas de Reconocimiento Automático del Habla.*** Apuntes de la UPM (Universidad Politécnica de Madrid), Facultad de Informática, Octubre de 2001.

[Rabiner1989] Lawrence R. Rabiner. ***A tutorial on hidden Markov models and selected applications in Speech Recognition.*** Proceedings of the IEEE, Vol. 77. No. 2, February 1989.

[Reynolds2000] Douglas A. Reynolds, Thomas F. Quatien and Robert B. Dunn. ***Speaker Verification Using Adapted Gaussian Mixture Models.*** MIT (Massachusetts Institute of Techonology) Laboratory. Digital Signal Processing, Vol. 10, Number 1-3, pages 19-41. January/April/July 2000.

[Esteve2007] C. Esteve Elizalde. ***Reconocimiento de Locutor Dependiente de texto mediante adaptación de Modelos Ocultos de Markov Fonéticos.***, Tesis de Licenciatura de la UAM (Universidad Autónoma de Madrid). Julio de 2007.

[Sarria2009] Milton Orlando Sarria Paja. ***Detección de Patologías en Señales de Voz mediante HMM empleando Entrenamiento Discriminativo.*** Tesis de Maestría de la UNC(Universidad Nacional de Colombia) Manizales, Colombia, Mayo de 2009.

[Rouas2005] Jean-Lue Rouas, Jerome Farinas, Francois Pellegrino, Regine Andre-Obrecht. ***Rhythmic unit extraction and modelling for automatic language identification.*** Speech Communication 47 (2005) 436456, april 26, 20005.

[Calvel2007] Chloé Calvel.***Analyse et reconnaissance des manifestations acoustiques des emotions de type peur en situations anormales.*** Tesis de Doctorado de la Éscole Nationale Supérieure des Télécomunicaciones. París, 15 de marzo de 2007.

[Rigaldie2004] Karine Rigaldie, Jean Luc Nespoulous, Nadine Vigouroux. ***Dysprosody in Parkinsons Disease: An Acoustic Study Based On Tonal Phonology and the INTSINT System.*** Dans : Speech Prosody 2004, Nara, Japan, 23/03/2004-26/03/2004, B. Bel, I. Marlien (Eds.), SProSIG, ISBN 2-9518233-1-2, p. 617-620, mars 2004.

[Kapoor2011] Tripti Kapoor, R. K. Sharma. ***Parkinsons disease Daignosis using Mel-Frecuency Cepstral Coefficients and Vector Quantization.*** International Journal of Computer Applications (0975-8887), Vol. 14-No.3, January 2011.

[Sánchez2009] Inocencio Sánchez Ciudad. ***Tecnología del Habla, Reconocimiento de Locutores y de Voz.*** Apuntes Curso 2009/2010 Tema5, de la

UCLM (Universidad de Castilla La Mancha), Escuela Superior de Ciudad Real. *http://www.inf-cr.uclm.es/www/isanchez/techabla0910/*

[Rabiner2010] Laurence R. Rabiner. ***Digital Speech Processing-Lecture 3: Acoustic Theory of Speech Production.*** Center of Advanced Information Processing of UCSB (University of California- Santa Barbara). Basic Course Material Fall 2010.

[Miyara2004] Federico Miyara. ***La voz humana.*** Laboratorio de Acústica y Electroacústica, Escuela de Ingeniería Electrónica, Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Argentina. Noviembre 2004. *http://www.fceia.unr.edu.ar/acustica/audio/index.htm*

[Godino2007] Juan Ignacio Godino Llorente. ***El diagnóstico y la evaluación de patologías de la voz a través de medidas no invasivas.*** Apuntes de la UPN (Universidad Politécnica de Madrid) Depto. ICS, Agosto de 2007.

[DelPino2004] Paulino del Pino, José A. Díaz, Carlos Jiménez, Howard B. Rothman. ***Identificación de algunos parámetros espectrales que determinan la calidad de la voz.*** Revista Ingeniería UC. Vol. 11, no 3, paginas 7-16, 2004.

[DelPino2008] Paulino Del Pino, Iván Granillo, Mario Miranda, Carlos Jiménez, José A. Díaz. ***Diseño de un sistema de medición de parámetros característicos y de calidad de señales de voz.*** Revista Ingeniería UC, vol. 15 no 2, páginas 13-20, 2008.

[Apollo] ***Apollo Hospitals***. *qualcommlife.com*

[GlucoHealth] ***MyGlucoHealth Wireless***. *qualcommlife.com*

[González1993] L. González Abril. ***Modelos de Clasificación basados en máquinas de vectores de soporte.*** Departamento de Economía Aplicada I, Universidad de Sevilla.

[Carvajal2010] Johanna Paola Carvajal González. ***Metodología de entrenamiento de modelos de mezclas Gaussianas empleando criterios de gran margen para la detección de patologías en bioseñales.*** Tesis de Maestría, Universidad Nacional de Colombia, Enero de 2010.

[Reynolds1995] Douglas A. Reynolds and Richard C. Rose. ***Robust text-independent speaker identification using gausian mixture speaker models.*** IEEE log number 9406779.

[Rabiner1993] Lawrence R. Rabiner, Bijng-Heang Juang. ***Fundamentals of Speech Recognition.*** ISBN 0130151572 9780130151575. Prentice Hall, 1993.

[DSP2000] ***Digital Signal Processing.*** A Review Journal. Academic Press, Vol 10, Numbers 1-3 January/April/July 2000.

[Galbiati] Jorge Galbiati R. ***Análisis Discriminante.***

[SPRO2009] ***SPRO: Speech Signal Processing Toolkit***, release 4.1. July 2009. *http://gforge.inria.fr/projects/spro/*

[PRAAT] **Praat.** *http://www.fon.hum.uva.nl/praat/*

[BECARS] ***BECARS. Programa para la Identificación de Locutor.*** *www.tsi.enst.fr/becars/index.php*

[Velázquez2011O] M. Edith Velázquez Vargas, Eduardo Sánchez Soto, Sergio Ivvan Valdez Peña, Eduardo Ortiz Hernández.***Modelado Estadístico del Estado Anormal de la Voz con Modelos Gráficos.*** 13^o Foro Estatal de Investigación e Innovación Oaxaca 2011, del 1 al 2 de diciembre, 2011. Instituto Tecnológico de Oaxaca, (ITO).

[Velázquez2011H] M. Edith Velázquez Vargas, Eduardo Sánchez Soto, Sergio Ivvan Valdez Peña. ***Modelado estadístico de características suprasegmentales y de frecuencia del estado anormal de la voz.*** 6^o Encuentro de Matemáticas Aplicadas a la Biología y Ciencias de la Computación del 22 al 25 de noviembre, 2011. Universidad Autónoma del Estado de Hidalgo, (UAEH).

[Sánchez2005] Eduardo Sánchez Soto. ***Réseaux Bayésiens Dynamiques pour la Vérifucation du Locuteur.*** Tesis de Doctorado de la ENST(École Nationale Supérieure des Télécommunications de Bretagne), Mayo 2005.

[Álvarez2005] Mauricio Álvarez, Germán Castellanos. **textitSelección de Características utilizando HMM para la identificación de Patologías en la Voz.** Scientia et Technica Ao XI No 28 Octubre de 2005 UTP (Universidad Tecnológica de Pereira). ISSN 0122-1701.

[Vignolo2008] Juan Vignolo Barchiesi. ***Introducción al Procesamiento de Señales.*** Ediciones Universitarias de Valparaiso, ISBN 978-956-17-0426-8, Chile 2008.