Caracterización Automática del Llanto de Bebé para su Estudio con Modelos de Clasificación

por

Erika Amaro Camargo

Tesis sometida como requisito parcial para obtener el grado de **Maestría en Ciencias** en el área de **Ciencias Computacionales** en el Instituto Nacional de Astrofísica, Óptica y Electrónica.

Supervisada por:

Dr. Carlos Alberto Reyes García
Investigador titular del INAOE

Abstract

As a part of a project that seeks to support early detection of pathologies in newborn babies, this thesis proposes a system of Automatic Infant Cry Recognition based on a characterization defined by the combination of acoustical features, which are obtained by different extraction techniques. Experiments were performed to recognize three types of cry: normal, pathological cry of hypo-acoustic (deaf) infants and asphyxia. The fact that the parameters have been derived from different spectral representation of the signal, suggests the possibility of raising different combinations of features to provide benefits to improve the representation of each type of crying, and consequently, increase the final recognition rate. In general, four characteristics extraction techniques were used: LPC (Linear Predictive Coding), MFCC (Mel Frequency Cepstral Coefficients), Intensity and Cochleograms. The original characteristic vectors were reduced through two methods like: LDA (Linear Discriminant Analysis), and a proposed method which is called, "Reduction by Statistics Operations". The combination of characteristics was carried out using the reduced characteristic vectors. The use of cochleograms to classify infant cry is one of the contributions of this thesis work. According to experiments, it was observed that cochleograms equalized, and in some cases improved the results obtained by techniques such as LPC or MFCC, which are widely used in speech recognition for their good results. Several tests were performed to validate the characterization. By applying traditional techniques such as ten-fold-cross-validation, results of an accuracy of 98.66% were achieved with vectors formed by the combination of four types of features. Other tests, which we call “individual tests” achieved results of 100% for the classification of the deaf class. Finally we defined a knowledge base for the classification of baby's cry considering the results and observations derived from this research.

Resumen

Como parte de un proyecto que busca apoyar la detección temprana de patologías en bebés recién nacidos, este trabajo de tesis propone un sistema de reconocimiento automático del llanto de bebés basado en una caracterización definida por la combinación de características, las cuales, son obtenidas por diferentes técnicas de extracción. Los experimentos se realizaron para reconocer tres tipos de llanto: normal (normo-oyente), patológico de bebés con hipoacusia (sordera) y asfixia. El hecho de que los parámetros hayan sido derivados de distintas representaciones de la señal, sugiere la posibilidad de plantear distintas combinaciones de características que aporten ventajas para mejorar la representación de cada tipo de llanto, y en consecuencia, aumentar la tasa de reconocimiento final. En general, se utilizaron cuatro técnicas de extracción de características: LPC (Codificación Predictiva Lineal), MFCC (Coeficientes Cepstrales de Frecuencia Mel), Intensidad y Cocleograma. Los vectores característicos originales fueron reducidos a través de dos métodos que son: LDA (Análisis Discriminante Lineal) y un método que se propone, el cual se denominó, “Reducción por Operaciones Estadísticas”. La combinación de características se llevó a cabo utilizando los vectores característicos reducidos. El uso de cocleogramas para llevar a cabo el reconocimiento automático del llanto de bebés es una de las aportaciones de este trabajo de tesis. De acuerdo a los experimentos realizados, se observó que los cocleogramas igualaron y en algunos casos mejoraron los resultados obtenidos por técnicas como LPC o MFCC, las cuales, son ampliamente utilizadas en reconocimiento de habla y han dado buenos resultados en reconocimiento de llanto. Diversas pruebas fueron realizadas para validar la caracterización. Aplicando técnicas tradicionales como validación cruzada, se lograron resultados del 98.66% de precisión con vectores formados por la combinación de cuatro tipos de características. Otro tipo de pruebas, las cuales denominamos pruebas por individuo arrojaron resultados de clasificación del 100% para la clase sordos. Finalmente, se define una base de conocimiento para la clasificación del llanto de bebé considerando los resultados y las aportaciones derivadas de este trabajo de investigación.

Dedicatoria

Dedico este trabajo de tesis a mi mamá, con amor, respeto y admiración, agradeciéndole todo el amor que siempre me ha mostrado.

Agradecimientos

Antes que todo quiero agradecer a mi asesor el Dr. Carlos Alberto Reyes García, por el apoyo, guía y consejos para el desarrollo de esta tesis, por darme la libertad de aportar mis ideas y dedicarme el tiempo siempre que necesité de su ayuda.

A mis padres, David Amaro y Antonia Camargo, por sus consejos, cuidados y amor incondicional, por creer en mí y darme la oportunidad de crecer libre.

A mis hermanos, David, Josué, Edna, Omar y Elizabeth, por ser la mejor familia para mí y apoyarme siempre.

A Javier Vázquez Cuchillo, por brindarme tu amor y confianza, por ser mi tranquilidad en la tormenta, gracias por tus palabras y por estar siempre a mi lado.

A José Alberto Méndez Polanco, por tus consejos, tu apoyo y tu confianza, pero sobre todo…por tu amistad.

A Javier Herrera Vega, por que aunque lejos, siempre has estado presente, gracias por ser mi amigo, te quiero mucho.

A Patricia Orta y Gustavo Hernández, por compartir conmigo su tiempo y brindarme su amistad.

A mis amiguis del cubo: Nadia Araujo, Coral Galindo, Rosario Peralta y Rosa María Ortega por tantos momentos compartidos y por su amistad.

A Carlos Guillén Galván por tu entrega y dedicación para ayudar a otros, mi agradecimiento y admiración para ti.

A la familia Vázquez Cuchillo, por abrirme las puertas de su hogar y darme su apoyo durante mi formación profesional.

De manera especial quiero agradecer a las personas que de alguna manera me brindaron su apoyo para salir adelante en esta etapa de mi vida, con la misma importancia agradezco:

A mis sinodales, Dr. René Armando Cumplido Parra, Dr. Jesús Ariel Carrasco Ochoa y Dr. Luis Villaseñor Pineda, por su paciencia y aportaciones para mejorar este trabajo de tesis.

Al Dr. Leopoldo Altamirano Robles por impulsarme a la superación constante.

Al Dr. Eduardo F. Morales Manzanares, por su confianza y apoyo.

Finalmente agradezco:

Al Instituto Nacional de Astrofísica Óptica y Electrónica por su apoyo durante la realización de mis estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) de México, por el apoyo económico proporcionado para mis estudios de maestría, bajo el número de beca 201668.

Este trabajo es parte de un proyecto
financiado por CONACYT
número 46753.

Contenido

PREFACIO

La detección temprana y no invasiva de patologías es un campo que actualmente se encuentra en desarrollo. Diversas investigaciones se han realizado con el objetivo de detectar patologías que afectan al ser humano y en las cuales una detección temprana permite la intervención oportuna, lo cual, reduce los efectos negativos que llegan a causar. Conjuntamente, el avance de tecnologías realizadas por computadora ha permitido que los resultados en esta área sean cada día más confiables y precisos. Cuando una persona padece de algún dolor en su cuerpo, o tiene molestias de cierto tipo, lo comunica al médico, y éste, con base en su conocimiento y experiencia, puede diagnosticar y definir la enfermedad del paciente. En el caso de un bebé, el único medio que tiene para comunicarse es a través de su llanto. El llanto como una herramienta útil para diagnosticar patologías relacionadas con el sistema nervioso central ha sido estudiada desde los 60's [1] [2]. En este trabajo el problema se aborda para la detección de bebés con hipoacusia, actualmente dicha detección en el campo médico se lleva a cabo mediante estudios directos sobre el infante. La idea de un dispositivo capaz de diagnosticar automáticamente ciertos problemas en un bebé, sin ser invasivo o sin tener que esperar resultados de laboratorio es muy atractiva para los científicos y médicos. Sobre todo para ser aplicada en aquellas zonas en las que no se cuenta con médicos especialistas.

El control de la audición del niño en todas las etapas de su vida es fundamental, sobre todo en los primeros años. La hipoacusia es una patología que consiste en la disminución o pérdida de la audición de uno o ambos oídos, lo cual impacta negativamente en el desarrollo del lenguaje y también en el desarrollo cognoscitivo y social del niño que la padece. Resulta entonces imprescindible el diagnóstico temprano de modo que se de inicio, lo mas pronto posible, al tratamiento e intervenciones necesarias para alcanzar un mejor desarrollo, principalmente en el área de la comunicación del infante.

Las estadísticas revelan que la hipoacusia de moderada a severa entre recién nacidos es de 1-3/1000. En el caso de la hipoacusia moderada o grave, el 50% de la misma es de origen genético. Algunas hipoacusias genéticas se acompañan de signos que orientan hacia un síndrome determinado. Pero la gran mayoría de las hipoacusias de origen genético no forman parte de un síndrome, por lo cual los niños afectados se presentan en apariencia normales pero con problemas serios en su audición [3]. De acuerdo con estudios estadísticos [4] y con cifras de la Organización Mundial de la Salud [39], una de cada 10 personas tiene algún tipo o grado de problema auditivo y dos de cada mil, presentan pérdida auditiva profunda o sordera bilateral. Esto implica que en México, hay 10 millones de personas con problemas de audición y 200,000 tienen sordera. Además, por estos datos y por el ritmo de crecimiento poblacional, se puede afirmar que en México, cada año, hay 4000 nuevos sordos.

Con un diagnóstico oportuno, muchos problemas de audición pueden y deben prevenirse; otros, pueden corregirse con tratamiento médico o quirúrgico y muchos más pueden ser compensados eficazmente con la adaptación profesional y cuidadosa de auxiliares auditivos, o inclusive, para casos con grados muy profundos de hipoacusia considerar los implantes cocleares. Debe recordarse que las sorderas prelingüísticas, además de la discapacidad auditiva ocasionan la incapacidad de desarrollo del propio lenguaje expresivo, y además bloquean la capacidad de adquirir el segundo gran código comunicativo lingüístico que se basa en la lectura y la escritura [5].

Este trabajo de investigación aborda el problema que existe para diagnosticar oportunamente la hipoacusia en bebés. Partiendo de grabaciones de llantos hechas por médicos especialistas en la detección de dicha patología, la señal es procesada aplicando técnicas de extracción de características de la señal, posteriormente, haciendo uso de métodos de reducción de datos se elimina información redundante, y finalmente, se utiliza un clasificador para determinar el tipo de llanto, ya sea llanto patológico o llanto de un bebé sano.

Para lograr los objetivos, este trabajo se centra en cuatro aspectos: primero, extraer diferentes características acústicas de la señal de llanto, explorando el uso de cocleogramas para caracterizar el llanto de bebés hipoacúsicos; segundo, aplicar métodos de reducción de datos sobre los vectores característicos extraídos; tercero, proponer una mejora a la representación del tipo de llanto mediante la combinación de diferentes tipos de características en un solo vector; y cuarto, iniciar la construcción de una base de conocimiento para la clasificación de diferentes tipos de llanto: normal, sordo y asfixia.

Organización de la tesis

Este trabajo de tesis se encuentra organizado en seis capítulos. En el capítulo 1 se define el estado del arte, los objetivos y la metodología de esta tesis. El capítulo 2 describe la importancia que tiene el llanto de bebé como herramienta para detectar la hipoacusia y se presentan los fundamentos teóricos para el procesamiento de la señal de llanto y su clasificación. En el capítulo 3 se realiza un estudio de las técnicas empleadas para extraer características de la señal y que son la base para llevar a cabo el reconocimiento automático. En el capítulo 4 se describe el proceso de reconocimiento automático de llanto de bebé que se propone. En el capítulo 5 se dan a conocer los experimentos y la evaluación realizada al sistema propuesto y finalmente, en el capítulo 6 se presentan las conclusiones y se plantean las perspectivas de esta tesis.

Capítulo 1

INTRODUCCIÓN

En este capítulo se presenta el estado del arte en el área de reconocimiento automático del llanto de bebé. Se definen los objetivos de la tesis, la metodología a seguir para alcanzar dichos objetivos, así como los alcances y limitaciones de la misma.

1.1 Antecedentes y estado del arte

Con el avance de la tecnología, los investigadores han sido capaces de precisar la estimación de características del sonido, particularmente, características de la señal de voz. En este trabajo, la señal de llanto de bebé es tratada como habla, por lo que las técnicas aplicadas para extraer características de la señal, son las mismas que se aplican para reconocimiento de habla o voz.

El análisis de llanto infantil en el área médica reportado en la literatura está basado principalmente en el análisis de espectrogramas. Los métodos realizados de esta manera son fuertemente dependientes de una evaluación subjetiva y no son convenientes para el uso clínico. En un trabajo de investigación realizado recientemente [6] se concluyó que "Es difícil establecer si el análisis espectro-fonográfico del llanto puede utilizarse como indicador temprano de alteraciones neurológicas. Los resultados podrán usarse para establecer comparaciones con otras enfermedades. Es necesario estandarizar el método óptimo de análisis del llanto". En otro estudio [7] se llegó a la conclusión de que "el análisis del espectrograma no tuvo la capacidad de identificar la hipoacusia en niños de corta edad, por lo que creemos que no es útil como método de detección temprana de sordera". Por lo tanto, se hace necesaria la propuesta de nuevos métodos o sistemas que ayuden a reforzar la detección temprana de este tipo de enfermedades en los bebés.

Dentro de los esfuerzos realizados para investigar el significado del llanto de bebé, se han desarrollado trabajos que a través de diversas técnicas, cada uno ha tenido logros importantes. En 1984 Cohen y Zmora [8], propusieron un sistema de clasificación basado en Modelos Ocultos de Markov de Densidad Continua, el cual fue probado con una base de datos constituida por llantos de hambre y dolor de bebés totalmente sanos, utilizando 10 LPC (*Linear Predictive Coding*) obtuvieron un 90% de clasificación correcta [16]. Petroni et al. en 1995 [9] utilizaron varios tipos de Redes Neuronales para clasificar llanto de bebé cuyas clases fueron dolor y no-dolor, sus resultados obtuvieron hasta un 86.2% de precisión. En el 2002 Taco Ekkel [10] utilizando un clasificador basado en una Red Neuronal de Base Radial clasificó sonido de llanto de bebés recién nacidos en dos tipos que llamó normal y anormal (hipoxia[1]), reportando resultados de clasificación correcta alrededor de 85%. También en el 2002 Lederman [11] propuso un sistema de reconocimiento de llanto basado en Modelos Ocultos de Markov, en el cual se clasificaron diversas patologías como: síndrome de dificultad respiratoria con un 63% de precisión, paladar hendido considerando varios escenarios logrando resultados desde 57.3%-90% de clasificación correcta, clasificó también llanto de bebés expuestos al consumo de drogas como cocaína y opio obteniendo un 46 y 63% de clasificación correcta respectivamente. Entre los trabajos más recientes dedicados al estudio de extracción de características y clasificación del llanto de bebés hipoacúsicos y normo-oyentes, se encuentra el trabajo de Orozco en el 2003 [12], donde se analizaron distintas características acústicas obtenidas de la señal. Orozco analizó cuatro tipos de características que fueron, Coeficientes de Predicción Lineal (LPC), Coeficientes Cepstrales de Frecuencia Mel (MFCC), Frecuencia Fundamental (Pitch) e Intensidad. Cada una de estas características fue obtenida mediante distintas técnicas y fueron probadas independientemente una de otra utilizando un clasificador basado en Redes Neuronales. Los mejores resultados se obtuvieron usando las características extraídas mediante la técnica llamada Codificación Predictiva Lineal o análisis LPC, la cual es

[1] Hipoxia: disminución del nivel de oxígeno en la sangre o en los tejidos (un tipo de asfixia), más adelante se dará una definición más detallada.

muy utilizada en el área de reconocimiento de habla. Sin embargo, debido al número limitado de muestras en ese momento, algunos ejemplos fueron duplicados para el entrenamiento de la red neuronal, lo cual representa un sesgo al momento de obtener los resultados que llegaron hasta un 98% de clasificación correcta. En el 2004 Réyes-Galaviz et al. [32] realizaron experimentos utilizando un modelo híbrido de Red Neuronal con Lógica Difusa (ANFIS) para clasificar tres tipos de llanto, entre ellos, sordera y asfixia obteniendo resultados del 93 al 96% de precisión. También en el 2004, utilizando características extraídas por LPC y como clasificador una red neuronal *Feed Forward Input Delay*, Réyes-Galaviz clasificó llanto patológico de sordera y asfixia, utilizando vectores sin reducción y una base de llantos de 6 bebés con sordera, 7 de asfixia y 85 normales, obtuvo un 98.67% de clasificación correcta [33]. En el 2006 Barajas [16] presentó diversos resultados clasificando llanto de dolor y hambre principalmente, realizando además pruebas para clasificar niveles de sordera. En todas las pruebas utilizó las características obtenidas por las técnicas MFCC (*Mel Frecuency Cepstrum Coefficients*) y LPC obteniendo los mejores resultados usando MFCC con precisión de hasta un 95 y 96% de clasificación correcta para hambre y dolor respectivamente. En Santiago de Cuba, el grupo de procesamiento de voz de la Universidad de Oriente, dirigido por Sergio D. Cano, ha investigado el llanto de bebé desde hace varios años [13]; sus resultados más recientes se presentan en el 2007 [14] donde se estudia el llanto de bebé para detectar hipoxia, mostrando una precisión de hasta un 85% de clasificación correcta. El uso de cocleogramas para caracterizar el llanto de bebé es una de las propuestas de este trabajo de tesis. Los cocleogramas han sido utilizados para reconocimiento de habla con buenos resultados. En el trabajo de Shamma, et al. [15] se usaron cocleogramas para el reconocimiento de habla a nivel de fonemas, superando a los resultados obtenidos con características obtenidas por LPC.

1.2 Objetivos

Objetivo general

Extraer y analizar características acústicas del llanto de bebés hipoacúsicos y normo-oyentes, incluyendo cocleogramas, con el fin de generar de manera automática una caracterización que permita su estudio con modelos de clasificación.

Objetivos particulares

- Extraer y analizar características acústicas de la onda de llanto de bebés hipoacúsicos y normo-oyentes incluyendo cocleogramas.
- Seleccionar y analizar técnicas de reducción de datos y explorar el uso de operaciones estadísticas para este fin.
- Proponer una mejora a la representación del tipo de llanto mediante la combinación de características en un solo vector.
- Seleccionar un modelo de clasificación que mejor se adapte a la caracterización generada.
- Iniciar la construcción de una base de conocimiento basada en los mejores resultados obtenidos, con el fin de servir como referencia futura para la clasificación de diferentes tipos de llanto: normal, sordo y asfixia.
- Evaluar y analizar la caracterización comparando los resultados del modelo de clasificación con trabajos anteriores.

1.3 Alcances y limitaciones

El reconocimiento automático del llanto de bebé es una línea de investigación abierta, la cual, en los últimos años, ha despertado un creciente interés por su utilidad como apoyo en el diagnóstico de patologías. En esta tesis se aborda el problema de clasificación de llanto patológico de bebés con hipoacusia (sordos). Se propone caracterizar la señal de llanto por medio de la combinación de características obtenidas por diversas técnicas: dos basadas en modelos perceptúales (MFCC y

cocleograma) y dos basadas en modelos de emisión de la voz (LPC e intensidad). Lo que se busca al combinar dichas características es mejorar la representación de cada tipo de llanto y, en consecuencia, mejorar la precisión de la clasificación que se obtiene probando cada técnica de manera independiente. En este trabajo, el problema de clasificación se aplica principalmente para dos tipos de llanto: normal y sordo; posteriormente se realizan pruebas para clasificar llanto de bebés con asfixia. Dado que se cuenta con pocas muestras de llanto de asfixia, sólo se realizan algunas pruebas para este tipo de llanto.

1.4 Metodología

Se propone la siguiente metodología para alcanzar los objetivos enunciados anteriormente:

1. **Extracción y análisis de características acústicas de la onda de llanto de bebés hipoacúsicos y normo-oyentes incluyendo cocleogramas.**

    - Se requiere una fase de preprocesamiento de la onda de llanto para unificar los parámetros y las propiedades de las muestras, en este caso realizar un remuestreo de las señales grabadas.
    - La extracción de características se aplica sobre segmentos de la señal, dicha segmentación y extracción se llevará a cabo utilizando el software PRAAT [25], este software permite trabajar con los sonidos del llanto, realizar análisis y manipular señales.
    - Se llevará a cabo un estudio de las técnicas de extracción de características, así como la obtención de cocleogramas de cada muestra de llanto. Al final del proceso de extracción se obtendrán las matrices de datos que contengan los vectores característicos.

2. **Selección y análisis de técnicas de reducción de información y exploración del uso de operaciones estadísticas para este fin.**

    - Existen varias alternativas para llevar a cabo la reducción de datos, entre las más utilizadas se encuentra la técnica de Análisis de Componentes Principales (PCA), y Análisis Discriminante Lineal (LDA).
    - Se propone la aplicación de operaciones estadísticas como el mínimo, máximo, promedio, desviación estándar y varianza, para llevar a cabo la reducción. Cada operador será aplicado a todo el conjunto de atributos de un vector, obteniendo sólo un valor representativo de los datos por cada operador.
    - Al final del proceso se obtendrán nuevos vectores característicos reducidos. Al menos dos métodos de reducción serán aplicados para realizar una comparación de los resultados.

3. **Selección de un modelo de clasificación que mejor se adapte a la caracterización generada.**

    - Realizar un análisis de los métodos de clasificación, seleccionando el más apropiado para los vectores característicos formados. Dentro de las alternativas se contemplan principalmente sistemas híbridos: con redes neuronales, lógica difusa, ensambles, máquinas de soporte vectorial, etc.
    - Probar el modelo de clasificación seleccionado. El lenguaje de programación dependerá de las características del clasificador que se seleccione.
    - Realizar pruebas usando los vectores característicos formados previamente.

4. **Proponer una mejora a la representación de los tipos de llanto mediante la combinación de características en un solo vector.**

    - En la combinación de características se usarán vectores de datos reducidos con el fin de acelerar el procesamiento. La evaluación de la mejora a la representación del llanto de bebé consistirá en verificar si efectivamente la combinación de características de distinta naturaleza aumenta la tasa de reconocimiento final para cada tipo de llanto.

5. **Evaluación y análisis de la caracterización haciendo una comparación de los resultados obtenidos.**

    - Evaluar la caracterización conforme a los resultados obtenidos por el modelo de clasificación. Las pruebas se realizarán de tres maneras, la primera, consistirá en evaluar el modelo de clasificación por medio de un conjunto de entrenamiento y un conjunto de prueba; el segundo tipo de pruebas será aplicando la técnica tradicional de validación cruzada; tanto en las primeras como en las segundas pruebas, se clasificarán segmentos de llanto de manera general, sin considerar a que muestra de llanto pertenecen, por lo que el tercer tipo de pruebas consistirá en separar los *n* segmentos que correspondan a una grabación completa de llanto, probar cada segmento en el modelo y, al final, realizar una votación en la cual se considerará correcta la clasificación, si más del 50% de los segmentos fueron clasificados correctamente.
    - Finalmente, hacer un análisis de los resultados mediante la comparación con trabajos previos tomando como parámetro la precisión del clasificador.

6. **Iniciar la construcción de una base de conocimiento basada en los mejores resultados obtenidos, con el fin de servir como referencia futura para la clasificación de diferentes tipos de llanto: normal, sordo y asfixia.**

    - Una vez terminadas las pruebas, se identificarán las mejores técnicas de extracción de características o combinaciones de ellas, para construir la base de conocimiento que sirva de referencia para futuras pruebas.

Capítulo 2

FUNDAMENTOS

2.1 El llanto de bebé y su importancia

Cuando nace un ser humano, éste depende completamente de la protección de los adultos, principalmente de su madre. Un recién nacido es incapaz de valerse por sí mismo durante un periodo relativamente largo y proporcionalmente largísimo con respecto a otros mamíferos. El único medio de comunicación con el que cuenta un bebé es su llanto, por medio del cual manifiesta sus diferentes estados, ya sea dolor, hambre, incomodidad, sueño, cansancio, etc. En el campo médico, los especialistas comentan que el llanto de bebé es una fuente valiosa de información para conocer si un infante tiene alguna alteración neurofisiológica, es decir, padece alguna patología en las funciones relacionadas con el sistema nervioso central. Lamentablemente, dichas patologías sólo pueden ser detectadas por personas especializadas, que a través del tiempo han desarrollado dicha capacidad de detección. Es importante diferenciar las diversas manifestaciones que tiene el llanto de un bebé y sobre todo darle la importancia como medio que tiene un bebé para comunicar que algo le sucede.

2.1.1 Hipoacusia neurosensorial infantil y asfixia - estadísticas

El diagnóstico temprano de la hipoacusia neurosensorial infantil no es tarea fácil. Su importancia reside en que si dicho déficit no es diagnosticado y tratado oportunamente en los primeros años de vida, genera alteraciones en el desarrollo lingüístico, intelectual y social del niño [17].

Es tal la preocupación mundial por este tema que aún sigue vigente la Asamblea realizada por la O.M.S. el 27 de marzo de l986 donde se llegó a la siguiente conclusión:

"…hasta el 5O% de los defectos de audición podrían evitarse o por lo menos disminuir sus secuelas por medio de la Prevención Primaria y Secundaria."

En México, un estudio realizado por el Instituto Nacional de la Comunicación Humana (INCH) de 1992 a 1996 en comunidades rurales de 16 estados de la República, comenta que: *los recursos (materiales y humanos) para combatir los defectos auditivos son casi inexistentes en nuestro país. La mayor parte de las personas con defectos auditivos viven en áreas marginadas (rurales o urbanas), donde es nulo el acceso a los servicios de audiología, medicina de la comunicación humana y otorrinolaringología* [4]. Desde el 2005 ya es obligatoria en México la detección de este mal. La capacitación se inició en ciudades del norte del país, como Tijuana, Monterrey, Chihuahua y Jalisco. Sin embargo, en muchos estados sólo el sistema de salud privado tiene la capacidad de detectar la sordera y atenderla.

En el caso de bebés recién nacidos que pasan por un período de la asfixia, éstos quedan expuestos a posibles cambios o alteraciones a nivel neurológico, en función del grado de asfixia que hayan sufrido. Según la Academia Americana de Pediatría (AAP), de 2 a 6 de cada 1000 recién nacidos de término completo presentan asfixia, y la incidencia es del 60% en recién nacidos prematuros con bajo peso. De ellos, alrededor del 20 al 50% mueren durante el período neonatal y de los sobrevivientes, el 25% desarrolla secuelas neurológicas permanentes [32].

2.1.2 Importancia de la detección temprana

Algunas sorderas son totalmente recuperables cuando se diagnostican y tratan a tiempo, éstas son las que se ocasionan por la infección del oído llamada otitis, en general se recuperan con tratamiento médico, pero en ocasiones, cuando se hacen crónicas deberán ser sometidas a tratamiento quirúrgico, antes de que se hagan definitivas. Otras no se recuperan a pesar del tratamiento y, cuando es así, un diagnóstico temprano modifica totalmente el curso de las complicaciones que tiene la

sordera en un niño. Cuando se tiene un diagnóstico completo y temprano (antes de los 6 meses), el niño tiene mejores oportunidades para su desarrollo integral.

Un niño sordo tiene grandes capacidades de atención visual, y sus horizontes de aprendizaje y desempeño profesional o tecnológico no deberían estar limitados si se maneja adecuadamente. Un niño que nace sordo y tiene intervención temprana puede concretar su proyecto de vida igual que cualquier otro niño sano [37].

La asfixia es otro tipo de problema que se presenta en bebés recién nacidos y es de vital importancia el detectarla, ya que la falta de una oxigenación adecuada puede causar daño en el cerebro, el corazón o los riñones, que son los órganos que sufren más la falta de oxígeno y en los que los daños suelen ser irreparables. Las vías respiratorias de un bebé que presenta asfixia pueden estar parcialmente bloqueadas. Una obstrucción parcial se puede convertir rápidamente en una situación potencialmente mortal si el bebé pierde la capacidad para inhalar y exhalar aire suficiente.

En términos médicos, la asfixia se cataloga en anoxia e hipoxia, los cuales, son dos fenómenos relacionados que se diferencian en el grado de severidad. Ambos se refieren a la capacidad del organismo de proveer oxígeno a los distintos tejidos del organismo, y por lo tanto, al contenido de ese gas vital. La diferencia entre anoxia e hipoxia es una cuestión de medida: cuando la cantidad de oxígeno disminuye por debajo del nivel de concentración normal el cuadro se denomina hipoxia, mientras que cuando el oxígeno está completamente ausente pasa a llamarse anoxia. En otras palabras: la anoxia es la ausencia de oxígeno que requieren los tejidos para mantener activo el ciclo celular [10].

2.1.3 Consecuencias de la sordera definitiva

Retraso o ausencia del lenguaje oral: a estos niños cuando no se atiende se les llama sordomudos, pero podrían no serlo si hubieran tenido la oportunidad de usar aparatos auditivos que les ayudaran a oír mejor y hubieran recibido rehabilitación para enseñarlos a hablar antes de los 5 años.

Problemas de aprendizaje y de relación con otros niños: generalmente son tímidos, tristes y en ocasiones son confundidos con niños que tienen problemas mentales.

2.1.4 Consecuencias de la asfixia

Todos los tejidos del organismo pueden verse afectados en forma a veces irreversible e irreparable por un cuadro de anoxia o por uno de hipoxia; dependiendo el daño del grado de dificultad en la llegada del oxígeno y en el tiempo que dure el cuadro. Es por esto último, que las distintas intervenciones médicas capaces de revertir estos cuadros de insuficiencia en el transporte de oxígeno, deben ser implementados lo más rápido que sea posible para evitar secuelas.

Pero sin lugar a dudas, el cerebro es el tejido neurológico que sufre con mayor severidad estos cuadros ya que, por su carácter inherente de ser irreparable, cualquier obstrucción en la llegada de oxígeno al mismo puede dejar secuelas permanentes. El daño que ocasiona la anoxia en el cerebro puede traducirse en la pérdida de aquellas funciones cognitivas, motoras o del lenguaje, cuyo sustrato orgánico se encuentra en la región cerebral afectada por el cuadro particular de falta de suministro de oxígeno. Otros tejidos que sufren en gran medida la falta de oxígeno son el corazón y el riñón [10].

2.2 Procesamiento Digital de Señales

El Procesamiento Digital de Señales o DSP's (por sus siglas en inglés) es un área que se dedica al análisis y procesamiento de señales (audio, voz, imágenes, video) que son discretas. Convierte señales de fuentes del mundo real (usualmente en forma analógica), en datos digitales que luego pueden ser analizados. Este análisis es realizado en forma digital, pues una vez que una señal ha sido reducida a valores numéricos discretos, sus componentes pueden ser aislados, analizados y reordenados más fácilmente que en su forma analógica [18].

La importancia que tiene la voz en el proceso de comunicación se ha incrementado con el rápido avance de la tecnología. La gran cantidad de posibilidades que la tecnología digital, basada en el desarrollo de microprocesadores cada vez más potentes, ofrece, ha hecho que las aplicaciones de procesamiento digital de señales se incrementen velozmente. Entre estas aplicaciones, las que tienen que ver con señales de voz han permitido el desarrollo de servicios que hasta hace algunos años eran impensables. Diálogo hombre-máquina, redes de integración de voz y datos, identificación y verificación de locutores, síntesis a partir de texto, son algunos ejemplos de los logros alcanzados por el procesamiento digital de señales de voz.

2.3 Codificación de la señal

El objetivo principal de la codificación de voz es la conversión de la señal de voz a una representación digital o secuencia binaria. Dado el carácter analógico (señal continua en tiempo y amplitud) de la señal de voz, la codificación de voz involucra un proceso básico de cuantificación y muestreo para obtener una representación digital (conversión analógico/digital). Mediante la cuantificación se discretiza la señal en amplitud y mediante el muestreo se discretiza la señal en tiempo. Para que en este proceso de digitalización exista el mínimo error de cuantificación, debemos muestrear la señal a una velocidad (frecuencia de muestreo) que como mínimo sea el doble de la

frecuencia más alta presente en la señal que estamos discretizando [19], a esto se le conoce como *Teorema de Nyquist.*

En el proceso de discretización en amplitud se debe utilizar un número de bits por muestra (*N*) que resulte adecuado para la calidad deseada. Así, por ejemplo, una señal de voz con calidad telefónica tiene una frecuencia máxima de 4 kHz lo que supone una frecuencia de muestreo mínima de 8 kHz y se suele utilizar una representación con 8 bits por muestra (256 niveles de cuantificación), lo que supone una velocidad de transmisión o necesidades de almacenamiento, en caso de grabación, de 64 kb/s.

Los codificadores de voz explotan las propiedades tanto temporales como frecuenciales de la señal de voz y del sistema auditivo humano, puesto que, en último término es el sistema auditivo humano quien juzga la calidad de la señal. El objetivo de la codificación digital de la señal de voz es representar la misma tan perfectamente como sea posible para que se pueda reconstruir una señal acústica a partir de la representación digital.

2.4 Producción de la señal de voz

Con el fin de comprender mejor el análisis de la voz y su representación digital es conveniente estudiar como se produce la señal de voz en el aparato fonador humano y el respiratorio.

En el habla humana, los sonidos se originan gracias al aire que sale de los pulmones y que llega al exterior a través de la laringe y del tracto vocal. En la laringe existen dos pliegues musculares (las cuerdas vocales) que al abrirse y cerrarse provocan la vibración de las moléculas del aire, emitiendo de este modo sonidos llamados *sonoros* como, por ejemplo, las vocales; en otras clases de sonidos, como en ciertas consonantes ([f] o [s], por ejemplo), el aire simplemente pasa entre las cuerdas vocales abiertas.

Este mecanismo permite dividir los sonidos en dos clases: los denominados *sonoros*, en los que vibran las cuerdas vocales, como se observa colocando las yemas de los dedos a la altura de la laringe y pronunciado una vocal alargada, y los *sordos*, en los que no se produce esta vibración (y que puede comprobarse por el mismo procedimiento, pronunciando una [s] larga).

Explicado de manera sencilla en [20], la señal de voz se produce por el funcionamiento secuenciado y sincronizado de los siguientes elementos:

1. Una corriente de aire proveniente de los pulmones y los músculos respiratorios.
2. Un vibrador sonoro constituido por las cuerdas vocales que están en la laringe.
3. Un resonador, conformado por la boca, la nariz y la garganta (o faringe).
4. Articuladores, labios, dientes, paladar duro, velo del paladar, mandíbula.

Estos cuatro elementos generan los sonidos del habla de la siguiente manera: primero, los pulmones suministran la corriente de aire que atravesando los bronquios y la traquea, sonorizan las cuerdas vocales que se encuentran en la laringe vibrando las mismas bajo la influencia del sistema nervioso. Es en la laringe donde propiamente se produce la voz en su tono fundamental y sus armónicos; luego sufre una modificación en la caja de resonancia de la nariz, la boca y la garganta (naso-buco-faríngea), que consiste en el aumento de la frecuencia de ciertos sonidos y la desvalorización de otros, formando el timbre de la voz y la calidad vocal, que son peculiares en cada persona. Los órganos articuladores (labios, dientes, paladar duro, velo del paladar, mandíbula) modelan finalmente el sonido. La Figura 2.4.1 muestra las partes del aparato fonador y respiratorio.

El funcionamiento de los músculos respiratorios, especialmente del diafragma, cambian en la pronunciación de distintos sonidos. Por otro lado, la altura de los sonidos verbales, depende de las oscilaciones de las cuerdas vocales, mientras que la

fuerza o intensidad depende de los cambios de presión de aire en la región de las cuerdas vocales, de la laringe y de la boca.

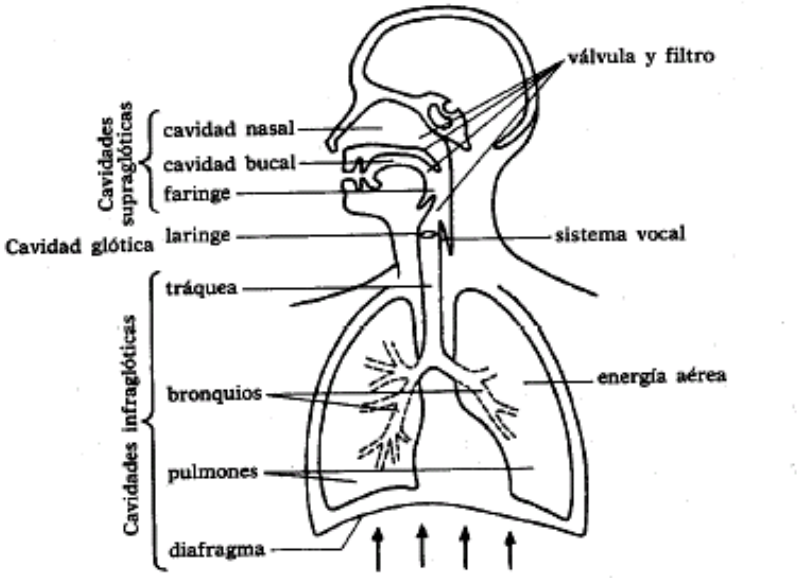


Figura 2.4.1. Conjunto de aparato fonador y respiratorio.
Imagen tomada de E. Martínez Celdrán (1984:76)

2.5 El sistema nervioso central y el oído

Algunos investigadores al referirse a la producción sonora, señalan que entran en funcionamiento de 90 a 100 músculos bajo el control del sistema nervioso central, donde cada músculo obedece a 14 órdenes por segundo. Así cuando se habla, el analizador motor capta los impulsos procedentes de los órganos del lenguaje a través de señales cinéticas [21]. Estos impulsos son componentes del segundo sistema de señales encargado de analizar, sintetizar y controlar la información a nivel cerebral, para seguidamente enviar órdenes a los efectores que van a poner en movimiento los órganos del habla.

En esta producción sonora, el oído desempeña un papel importante como regulador en el funcionamiento coordinado de los resonadores bucal-faríngeo. La pérdida

parcial o total de la audición altera dicho funcionamiento. El tono nasal del lenguaje de los sordos se debe, en parte, por ejemplo, a la falta de control auditivo en la regulación de los movimientos de la lengua y del analizador faríngeo [20].

2.6 Modelo digital de la producción de la señal de voz

Una aproximación razonable a la producción de voz desde el punto de vista de un modelo digital es la siguiente: se podría considerar a las cuerdas vocales como un generador de impulsos cuasi-periódicos que produce los sonidos sonoros. Por otra parte, los sonidos sordos tendrían su origen en un generador de números aleatorios. Según las características del periodo de la señal de voz en que nos encontrásemos se activaría uno u otro generador. La salida de ambos generadores pasaría a través de una cavidad resonante (cavidad bucal) considerada como un filtro digital variable en el tiempo. La amplitud de la señal de excitación a la entrada del filtro se regularía con un controlador de amplitud. Además, como las fuentes de sonido y la forma del tracto vocal son relativamente independientes, parece razonable modelarlas separadamente como se hace en el esquema del modelo digital de la Figura 2.6.1

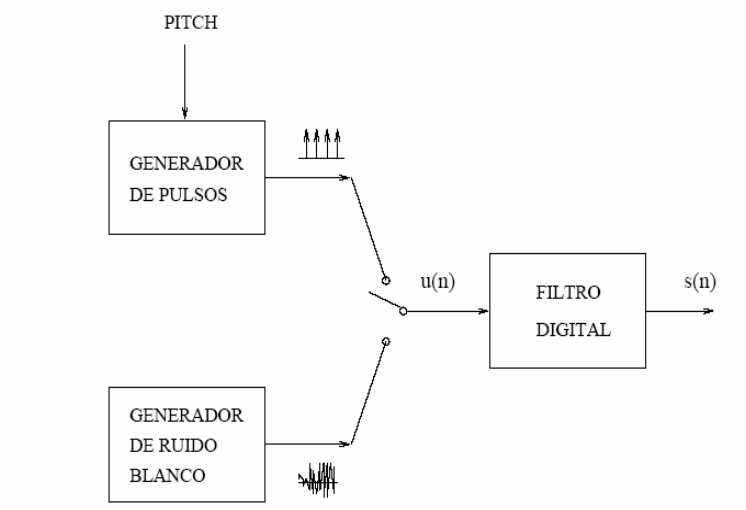


Figura 2.6.1. Modelo digital de producción de la señal de voz.

2.7 Análisis por intervalos cortos de tiempo (Short-Time Analysis)

Las propiedades de la señal varían con el tiempo. Sin embargo, si observamos dentro de ciertos bloques veremos que sus características permanecen esencialmente constantes. Debido a esta propiedad de las señales de voz, el análisis de las mismas suele hacerse tomando intervalos cortos de tiempo equiespaciados, de forma que las características de la señal dentro de ellos sean esencialmente invariantes. Analizando la señal de voz de ésta manera se alcanza una alta probabilidad de que el análisis sea eficaz y realmente descriptivo para poder utilizarlo posteriormente en la fase de reconocimiento.

2.8 Redes Neuronales Artificiales

Las redes neuronales artificiales son modelos que intentan reproducir el comportamiento del cerebro humano para adquirir el conocimiento. Se constituyen inicialmente como una simulación abstracta de los sistemas nerviosos biológicos formados por un conjunto de unidades llamadas neuronas o nodos conectados unos con otros. Una elección adecuada de sus características, más una estructura conveniente, es el procedimiento convencional utilizado para construir redes capaces de realizar una determinada tarea [22].

Las redes neuronales están compuestas de unidades denominadas neuronas, cada neurona tienen tres partes: una dendrita que recolecta las entradas desde otras neuronas o de un estímulo externo, un soma o cuerpo de la neurona que realiza un procesamiento no lineal sobre los estímulos de la entrada y finalmente un axón que transmite una señal de salida a otras neuronas. La conexión entre dos neuronas se denomina sinapsis [23].

Las neuronas se agrupan en capas, las neuronas conectadas unas a otras componen una red neuronal. Dependiendo de cómo estos componentes (neuronas en capas) son conectadas, diferentes arquitecturas pueden ser creadas (*Feed Forward Neural Network, Recurrent Neural Network*, etc.). La topología o arquitectura de la red neuronal consiste en el tipo, organización y disposición de las neuronas en la red, formando capas o agrupaciones de neuronas.

En este trabajo, el modelo de red neuronal que se utiliza es la red multicapa *feed-forwad* cuya imagen se muestra en la Figura 2.8.1. La topología de una red neuronal multicapa depende del número de variables en la capa de entrada, del número de capas ocultas de neuronas, del número de neuronas por cada capa oculta y del número de variables de salida en la última capa. Todos esos factores son importantes a la hora de determinar la configuración de una red neuronal. El término *feed-forwad* se refiere a que cada neurona está conectada sólo con las unidades de la siguiente capa, i.e. unidireccionales y sin ciclos.

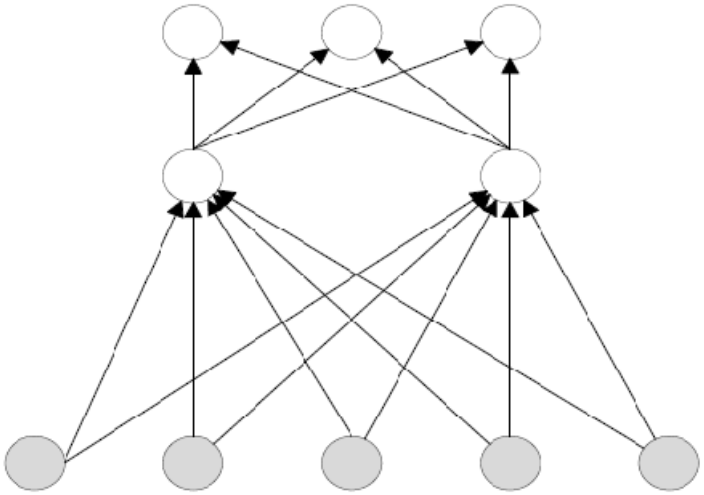


Figura 2.8.1. Red neuronal feed-forward
(5 entradas, una capa oculta con dos neuronas y 3 salidas)

2.8.1 Aprendizaje de una red neuronal

Una propiedad importante de las redes neuronales es la habilidad de aprender a partir de su ambiente. Eso es realizado a través de un proceso interactivo de ajustes aplicado a sus pesos de conexión entre dos neuronas, lo cual se denomina entrenamiento de la red. Para llevar a cabo este proceso, existen diferentes algoritmos y paradigmas que se aplican dependiendo de las características de la tarea a realizar, algunos de los más utilizados se muestran en la Figura 2.8.1.1.

Figura 2.8.1.1. Algoritmos y paradigmas de aprendizaje de una red neuronal.

En este trabajo, el paradigma que se utilizará será el de aprendizaje supervisado y el algoritmo de corrección de error.

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un metódico entrenamiento controlado por un agente externo, conocido como supervisor, que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y si ésta no coincide con lo que se quiere, se procede a modificar los pesos de las conexiones para conseguir que la salida obtenida se aproxime a la deseada.

El mecanismo más común de entrenamiento usado para redes multicapa es el algoritmo *Backpropagation*, en el cual, los pesos son actualizados por las capas

ocultas que adoptan el mecanismo de retropropagación para corregir la señal desde la capa oculta.

El aprendizaje de la red neuronal se puede especificar como una función de aproximación donde el objetivo es aprender una función desconocida (o una buena aproximación de la misma) desde un conjunto de pares de entrada.

Los factores que influyen en el proceso de aprendizaje del algoritmo *backpropagation* (BP) son, entre otros:

- Los pesos iniciales que son normalmente inicializados de forma aleatoria.
- La tasa de aprendizaje, que es un factor de gran importancia en el proceso de convergencia. A mayor tasa de aprendizaje, mayor es la modificación de los pesos en cada iteración, con lo que el aprendizaje es más rápido, pero por otro lado puede ocasionar oscilaciones en la convergencia.
- El momento (momentum), para controlar las oscilaciones se utiliza una constante de momento β, la cual determina el efecto en $t+1$ de los cambios de los pesos en el instante t. Con este momento se logra la convergencia de la red en menor número de iteraciones.
- Función de coste. En general, el entrenamiento de las redes neuronales se realiza a través de la minimización de una determinada función de coste. La función de coste que se más se emplea es el error cuadrático sobre las muestras de la función.
- Número de neuronas en las capas ocultas.

Básicamente el algoritmo de retropropagación se basa en la minimización de un error por medio de una técnica clásica de optimización llamada *descenso de gradiente*. Es decir, la idea principal consiste en calcular los pesos adecuados comparando la salida deseada con la respuesta de la red de manera que el error sea mínimo. La función que

usualmente es utilizada para medir el error es la suma de los errores al cuadrado o su promedio.

Capítulo 3

EXTRACCIÓN DE CARACTERÍSTICAS DE LA SEÑAL

3.1 Atributos acústicos

En este trabajo, el llanto de bebé es tratado como habla. Para seleccionar los métodos de extracción de características, se llevó a cabo un análisis de las técnicas más utilizadas en habla, y principalmente, en trabajos previos enfocados a la clasificación del llanto de bebé, de los cuales, tres métodos fueron seleccionados por sus buenos resultados para clasificar llanto [11][12][16]; estas técnicas son: LPC, MFCC e intensidad, a esta lista se adicionó el uso de cocleogramas, los cuales han mostrado buenos resultados en reconocimiento de habla [15], teniendo así cuatro técnicas de extracción de características. A continuación se describen cada una de ellas.

3.2 Codificación Predictiva Lineal (LPC)

La codificación predictiva lineal o LPC (*Linear Predictive Coding*), es una de las técnicas más utilizadas en codificación de voz, basada en la redundancia de la señal de habla en cuanto a periodicidad y variación relativamente lenta que permite la predicción de una señal muestreada a partir de muestras anteriores. Proporciona aproximaciones a los parámetros de la voz muy precisas utilizando la información de un modelo lineal. El modelo LPC está basado en el hecho de que una señal que transporta mensajes nunca es completamente aleatoria. Existe una correlación entre muestras sucesivas, y LPC utiliza esta correlación para reducir la cantidad de datos cuando se guarda la información de la señal. La popularidad de la técnica LPC se deriva de su precisa representación de la magnitud espectral de la voz, así como también de su relativa simplicidad de cálculo.

El método de predicción lineal recibe este nombre porque pretende extrapolar el valor de la siguiente muestra de voz *s(n)* como la suma ponderada de muestras pasadas *s(n-1), s(n-2), ...,s(n-k)*. Como se explicó, la señal de voz mantiene sus propiedades sustancialmente invariantes durante ciertos intervalos. El análisis LPC aprovecha esta característica basándose en que es posible determinar el valor de una muestra cualquiera de la señal a partir de las *p* últimas muestras. Es decir, las muestras de voz *s(n)* están relacionadas con la excitación *δ(n)*, se puede definir esta relación por medio de la siguiente ecuación:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + \delta(n)$$

donde *δ(n)* es la excitación que se aplica al filtro y puede consistir en impulsos periódicos (segmentos vocálicos) o en ruido aleatorio (segmentos no vocálicos), y los a_k son los parámetros de la función de transferencia de la cavidad bucal. La función de transferencia sin pérdida puede describirse como un modelo '*todo polos*'. Esta es también una aproximación razonable al lenguaje formado por la excitación del tracto vocal mediante pulsos de la glotis[2] (aunque los pulsos de la glotis no son espectralmente planos). La Figura 3.2.1 representa dicha función de transferencia.

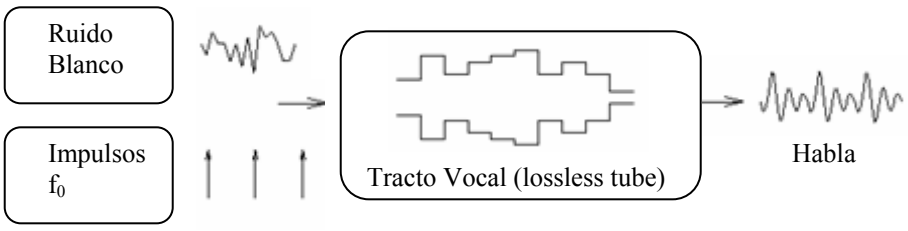


Figura 3.2.1. Representación de la función de transferencia de la cavidad bucal.

[2] Glotis: abertura anterior de la laringe. (ver Figura 2.3.1).

Realizando una aproximación por medio de un polinomio predictor lineal se obtiene la siguiente expresión:

$$\hat{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k)$$

donde los α_k son los coeficientes del predictor que hay que calcular. Una forma de hallar estos coeficientes es minimizando el error de predicción, es decir, minimizar la media de los cuadrados de la diferencia entre las muestras reales y las predichas linealmente durante un intervalo corto de tiempo. Dadas *p* muestras de lenguaje, se desea calcular estimaciones de α_k. Se debe realizar el cálculo de los coeficientes α_k minimizando alguna función de error *E*, concretamente de mínimos cuadrados, sobre una ventana de tamaño *N*.

El error en cualquier momento, e_n, es:

$$e_n = s(n) - \hat{s}(n)$$

$$= s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$

De aquí que la suma del error cuadrático, E, a lo largo de una ventana finita de longitud N, sea:

$$E = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} (s(n) - \sum_{k=1}^{p} \alpha_k s(n-k))^2$$

En general, el análisis predictivo lineal es una buena técnica de modelado para la señal de voz. Esto es especialmente cierto en las regiones de voz que más se aproximen a la casi estacionalidad, para las que el modelo '*todo polos'* proporcione una buena aproximación de la envolvente del tracto vocal. Durante las regiones de

transición y de voz sorda, este modelo es menos efectivo que para regiones de voz sonora, pero, aún así proporciona un modelo aceptablemente bueno para aplicaciones de reconocimiento del habla, aplicándose también para llanto de bebé.

3.3 Coeficientes Cepstrales de Frecuencia Mel (MFCC)

Los Coeficientes Cepstrales de Frecuencia Mel (MFCC), son coeficientes para la representación del habla basados en la percepción auditiva humana. Se derivan de la Transformada Discreta de Fourier (DFT) o de la Transformada de Coseno Discreta (DCT). Las bandas de frecuencia están situadas logarítmicamente (según la escala Mel), que modela la respuesta auditiva humana más apropiadamente que las bandas espaciadas linealmente de DFT o DCT. Adicionalmente, el espectro representado por los coeficientes, tiene una frecuencia similar a la del oído humano, la cual es más sensible a ciertas frecuencias que a otras. De esta manera lo que se obtiene es una aproximación a la forma en que el oído percibe los sonidos. Básicamente, los MFCC pueden obtenerse en los siguientes pasos:

1. Pre-énfasis.
2. Dividir la muestra en segmentos pequeños: Enventanado.
3. Calcular la DFT para cada segmento.
4. La escala de frecuencias se transforma a la escala MEL.
5. El espectro se convierte a la escala logarítmica.
6. Se calcula la transformada inversa discreta de Fourier.
7. En habla, típicamente se usan los primeros 13 coeficientes.

La Figura 3.3.1 muestra estos pasos, y a continuación se describen cada uno de ellos.

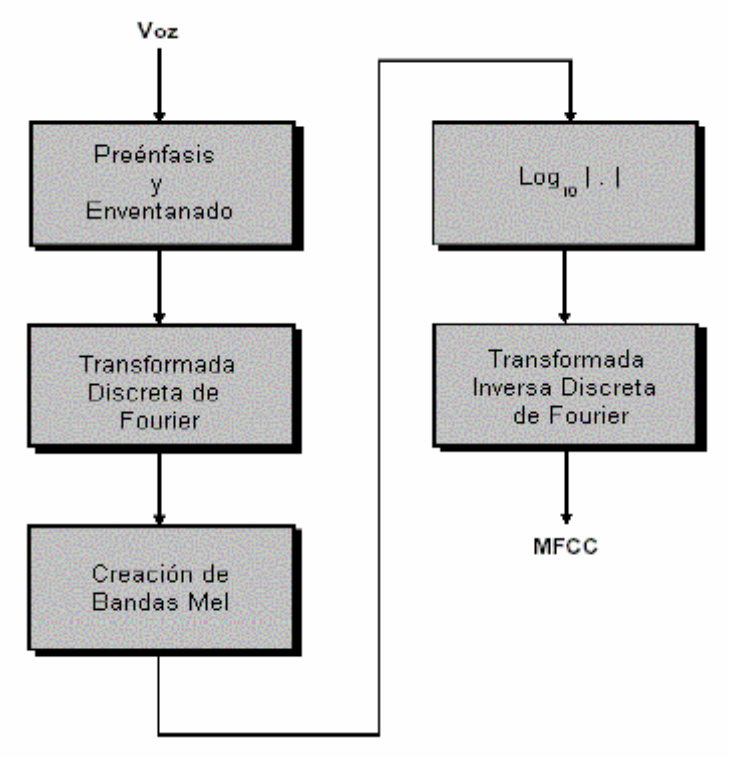


Figura 3.3.1. Esquema de obtención de MFCC

3.3.1 Pre-énfasis y enventanado (*Windowing*)

La inteligibilidad oral se debe a las altas frecuencias. Para que el habla sea comprensible, es indispensable la presencia de armónicos[3] cuya frecuencia se halla entre 500 y 3500 Hz. Por otra parte, la energía de la voz está contenida en su mayor parte en las bajas frecuencias y su supresión resta potencia a la voz que suena delgada y con poca energía. Debido a que la señal de voz se atenúa conforme aumenta la frecuencia, es necesario introducir un filtrado cuya función es incrementar la relevancia de las componentes de alta frecuencia. Este proceso se conoce con el nombre de pre-énfasis y puede ser diseñado a través de un filtro digital paso alto. Las

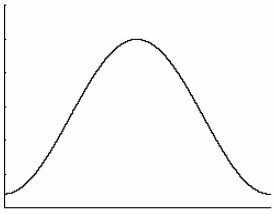
[3] En acústica, un armónico de una onda es un componente sinusoidal de una señal. Su frecuencia es un múltiplo de la fundamental. Los armónicos son los que generan el timbre característico de una fuente de sonido (ya sea una voz humana, un instrumento musical, etc.)

secciones de voz usualmente caen en las altas frecuencias debido a las características físicas del sistema de producción del habla. La mayoría del ruido está en las bajas frecuencias, eliminar estas bajas frecuencias y enfatizar las altas, mejora la audibilidad de la voz. Eliminar las bajas frecuencias hace que éstas resulten menos audibles, independientemente de que hayan sido generadas en la voz o en el ruido de fondo. Pero esto no es un problema fundamental porque el sistema auditivo humano es capaz de compensar estas bajas frecuencias que faltan en la señal.

Se aplica un ventaneado, que puede ser Rectángular o Hamming para seleccionar la trama con la que se va a trabajar. La ventana rectangular se define como:

$$Wn = \begin{cases} 1 & 0 \le n \le N \\ 0 & en\ otro\ caso \end{cases}$$

Sin embargo, la utilización de esta ventana trae consigo que en los puntos de inicio y fin exista una fuerte discontinuidad. Para reducir el efecto de discontinuidad al mínimo, se emplean tipos de ventanas que tiendan a reducir a 0 los valores de las muestras en los extremos. Aunque existen varios de tipos de ventana, la más común en análisis de voz es la ventana de Hamming (Figura 3.3.1.1) ya que tiene un efecto de estrechamiento en los bordes, mientras que no tiene efecto en la zona central lo cual atenúa la distorsión producida por las discontinuidades de los puntos de inicio y fin.



$$Wn = \begin{cases} 0.54 - 0.46 \cos(2\pi n/(N-1)) & 0 \le n \le N \\ 0 & en\ otro\ caso \end{cases}$$

Figura 3.3.1.1. Ventana de Hamming y su fórmula

3.3.2 Transformada Discreta de Fourier

Se aplica la transformada discreta de Fourier (DFT) sobre la trama de muestras ventaneada. La DFT (Discrete Fourier Transform) se calcula de la siguiente manera [24]: Tenemos una señal *x[n]* limitada a *N* muestras con un periodo de muestreo t_s de forma que $N \cdot ts = T$ (donde N es el número de muestras de la ventana que se va a analizar). Al calcular los coeficientes *X[k]* queda:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot \exp(-j2\pi kn / N) \quad k = 0, 1, 2, \ldots, N\text{-}1$$

La cantidad X[k] es la serie de Fourier Discreta de la señal periódica muestreada x[n].

La motivación del uso de la DFT parte del hecho de la utilidad que tiene para descomponer la señal de habla (llanto) en sus componentes en frecuencia. Un aspecto importante si queremos usar la DFT con señales de llanto es que debemos asumir que al menos en periodos cortos de tiempo se cumple que la señal es estacionaria. En la realidad esto no es estrictamente así aunque podemos suponerlo.

3.3.3 Aplicación de filtros Mel

Antes de ver como se aplican los filtros en la escala Mel, se presenta una breve descripción de dicha escala.

La escala Mel es una escala perceptual de frecuencias. El punto de referencia entre esta escala y la frecuencia normal se define equiparando un tono de 1000 Hz a 40dBs por encima del umbral de audición del oyente[4], con un tono de 1000 Mels.

[4] Umbral de audición = El umbral de audición define la mínima presión requerida para excitar el oído. Se ha tomado como convención, un umbral de audición de 0 dB que equivale a un sonido con una presión de 20 micropascales = 0.000002 pascales para frecuencias entre 2KHz y 4KHz.

Por encima de 500 Hz, los intervalos de frecuencia espaciados exponencialmente son percibidos como si estuvieran espaciados linealmente (Figura 3.3.3.1). La escala Mel resulta de dividir el espectro de frecuencias en un banco de filtros, mucho más estrechos y linealmente espaciados en las bajas frecuencias y, muy amplios y logarítmicamente espaciados en las altas. De este modo, se da mayor importancia a la información contenida en las bajas frecuencias, de acuerdo a la conocida variación de los anchos de banda del oído humano, y para capturar mejor las características fonéticas del habla.

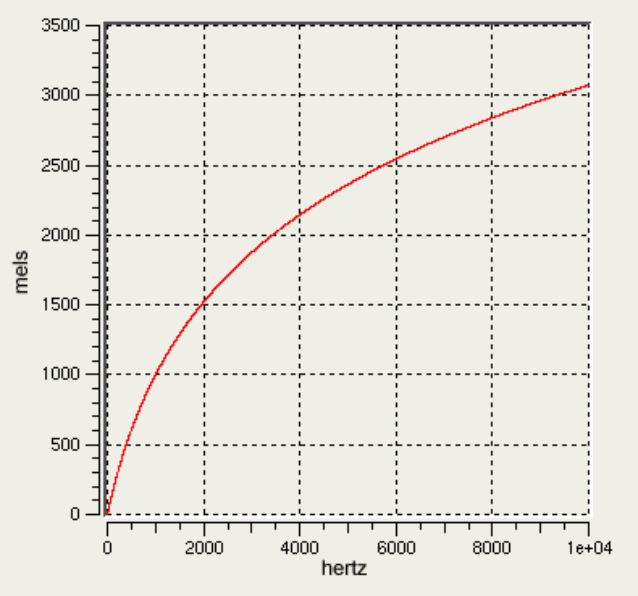


Figura 3.3.3.1. Escala Mel

Se calcula la energía en cada una de las bandas de frecuencias en que la escala Mel divide el espectro. Para ello se suman los módulos al cuadrado de la DFT en los puntos que se encuentren contenidos en cada una de dichas bandas. La Figura 3.3.3.2 muestra un ejemplo de este proceso.

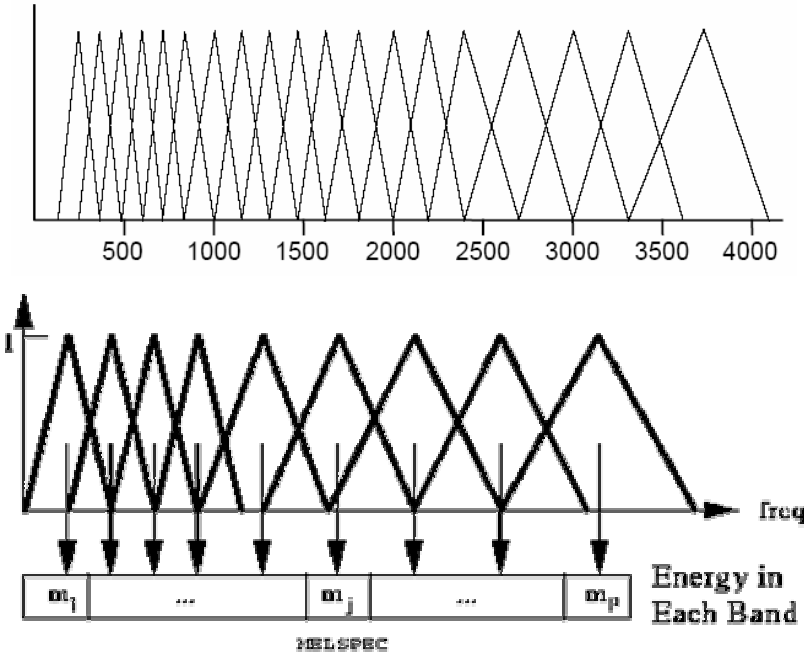


Figura 3.3.3.2. Banco de filtros en escala Mel

Cada magnitud de los coeficientes de DFT es multiplicada por la ganancia correspondiente del filtro relacionado. Uniformemente espaciado por debajo de 1kHz. y logarítmicamente escalado sobre 1kHz.

3.3.4 Aplicación de log

El logaritmo comprime el rango dinámico de valores, lo cual es una característica del sistema auditivo humano. El logaritmo también hace la extracción de características menos sensitiva a variaciones de la entrada.

3.3.5 Aplicar DCT

Los coeficientes cepstrales se calculan como la transformada coseno discreta (DCT), que hace las veces de transformada inversa de las energías logarítmicas obtenidas con anterioridad. En concreto, los coeficientes cepstrales se obtienen a partir del muestreo en *M* puntos de dicha transformada. El cálculo de los MFCC responde a la expresión:

$$MFCC_j(i) = \sum_{k=1}^{L} e(j,k)\cos\left[i(k-\frac{1}{2})\frac{\pi}{L} \right] \text{ donde } i=1\ldots M$$

donde, k es la banda de frecuencias, j es la trama en curso, $e(j, k)$ es el logaritmo de la suma de los módulos al cuadrado de la DFT en la banda k de la trama j, L es el número de bandas o filtros, M es el número total de coeficientes MFCC.

3.4 Intensidad

La intensidad se define como la amplitud de la onda sonora. Muchos sonidos presentan un patrón claro de intensidad que varía con el tiempo. La intensidad es la energía con la que el aire es impulsado desde los pulmones hacia las cuerdas vocales, de ésta forma, si hablamos en voz baja, la intensidad es muy débil, mientras que si hablamos en voz alta la intensidad será mayor y necesitaremos respirar con mayor frecuencia. La intensidad de la voz depende básicamente de la potencia con la que el aire que procede de los pulmones cuando hablamos golpea los bordes de la glotis, de modo que cuanto más amplias son las vibraciones que se producen durante la fonación, tanto mayor es la fuerza a la que se emite una voz. La unidad de medida de la intensidad es el Bel, aunque en la práctica se usa el Decibelio o Decibel (dB), que es una décima parte del Bel. Para tener idea, en una conversación normal, la intensidad de voz suele situarse en torno a los 50 dB. Sobre la intensidad de la voz, resaltaremos su capacidad para expresar también actitudes emocionales. De hecho, las variaciones de intensidad son muy adecuadas para representar estados de ánimo. Por otra parte, el tratamiento técnico de la intensidad, hace referencia a la manipulación de la amplitud. En el caso de un bebé sordo, la intensidad se ve afectada ya que no existe una retroalimentación del sonido y por lo tanto puede ser considerada como un parámetro importante para diferenciar entre un bebé sano y otro que no lo está.

3.5 Cocleograma

Un cocleograma representa los patrones de excitación de la membrana basilar (en el oído interno). Se basa en un modelo perceptual que utiliza la escala Bark (Figura 3.5.1) la cual es una escala psicoacústica[5]. La escala tiene un rango del 1 al 24 y corresponde a las primeras 24 bandas críticas del oído. Un *Barkfilter* tiene una escala de frecuencia que es altamente lineal debajo de los 1000Hz y fuertemente logarítmica sobre los 1000 Hz.

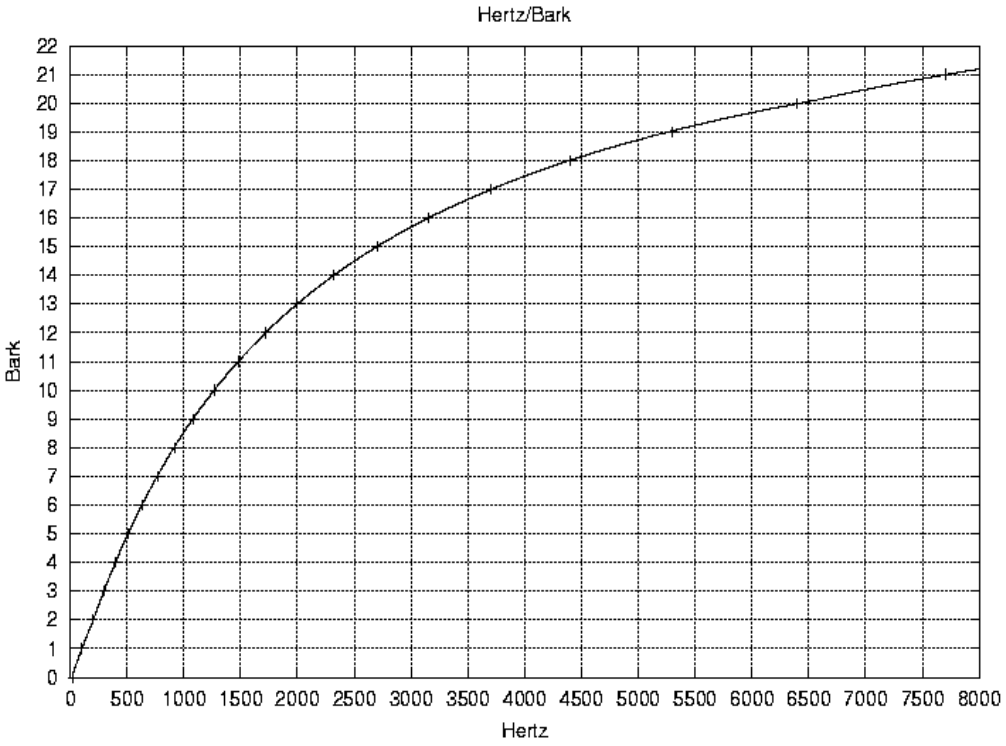


Figura 3.5.1 Escala Bark.

En un *Barkfilter* los logaritmos de las intensidades son mapeados. El cocleograma está basado en el *Barkfilter,* cuyo comportamiento es similar a la percepción del oído humano. El cocleograma usa la misma escala de frecuencia que el *Barkfilter* pero en el cocleograma en lugar de las intensidades lo que se mapea son los volúmenes de la

[5] La psicoacústica estudia la percepción subjetiva de las características del sonido: intensidad, tono y timbre. Estas características del sonido están, a su vez, determinadas por los propios parámetros del sonido, principalmente, frecuencia y amplitud.

señal según la percepción del oído. En el programa PRAAT [25], para una frecuencia dada en Hertz, la frecuencia correspondiente en Bark es encontrada con la siguiente fórmula:

$$Bark = 7x\ln\left(\frac{Hertz}{650} + \sqrt{1+\left(\frac{Hertz}{650}\right)^2}\right)$$

En la escala Hertz, el rango de frecuencia de percepción del ser humano es de 20 a 20,000 Hz que corresponden con un rango de frecuencia de 0.22 a 28.84 Bark.

En una *Barkfilter* para cada tiempo y por cada frecuencia la intensidad es dada. En un cocleograma para cada tiempo y por cada frecuencia se da el volumen. Cuando dos sonidos tienen la misma intensidad, pero diferentes frecuencias, probablemente serán percibidos con diferentes volúmenes. El volumen es una percepción de la intensidad y se expresa en referencia a intensidades. En un cocleograma el valor de referencia para la intensidad es 1000 Hz. La relación entre el volumen de referencia y el volumen de otra intensidad dada a una frecuencia específica es determinada experimentalmente.

Cuando un tono conduce a la activación de las células ciliadas (receptores auditivos) sobre una gran superficie en la membrana basilar, el oído no es capaz de percibir otras frecuencias vecinas y se dice que un tono es enmascarado por el otro. Hay dos tipos de enmascaramiento: lateral y hacia delante.

Enmascaramiento frecuencial o Lateral

Se produce cuando, al mismo tiempo, diferentes frecuencias vecinas se registran. Un tono puede hacer que otro tono sea casi inaudible. En general un tono bajo enmascara a un tono alto en lugar de lo contrario. Este fenómeno perceptivo puede explicarse de manera simplificada considerando como varía la excitación de la membrana basilar del oído según la frecuencia. Esta membrana vibra, en función de la tonalidad, más cerca o más lejos de la ventana oval. Un tono grave enmascara a uno agudo con más facilidad.

El enmascaramiento hacia delante

Aparece cuando un tono se produce después de otro. Por ejemplo, después de haber oído un fuerte sonido nuestros oídos quedarán saturados por un corto tiempo y los sonidos sucesivos se confundirán entre sí, y el sonido más fuerte será el que oculte a los otros.

En un cocleograma tanto el enmascaramiento lateral como hacia delante es modelado. La Figura 3.5.2 muestra el ejemplo de un cocleograma en su representación espectral para el llanto de un bebé normal, y en la Figura 3.5.3 para el llanto de un bebé sordo.

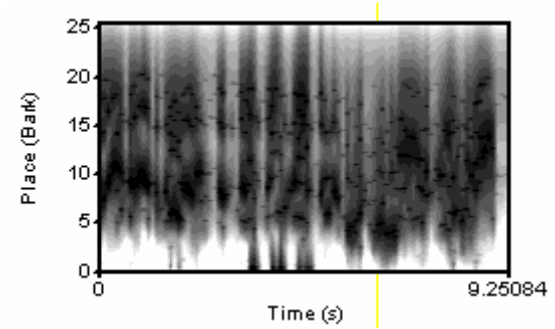


Figura 3.5.2 Cocleograma del llanto de un bebé normal

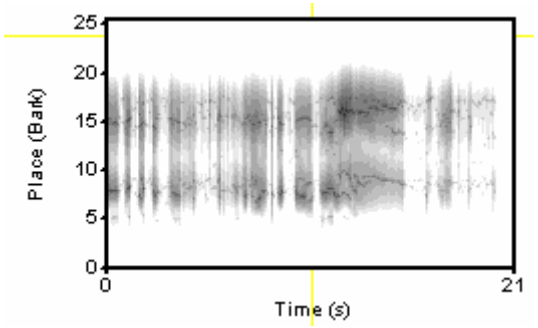


Figura 3.5.3 Cocleograma del llanto de un bebé sordo.

Capítulo 4

PROCESO DE RECONOCIMIENTO AUTOMÁTICO DEL LLANTO DE BEBÉ

4.1 Modelo global del sistema

En la Figura 4.1.1 se observan los módulos que integran el sistema de reconocimiento del llanto de bebé desarrollado en esta tesis.

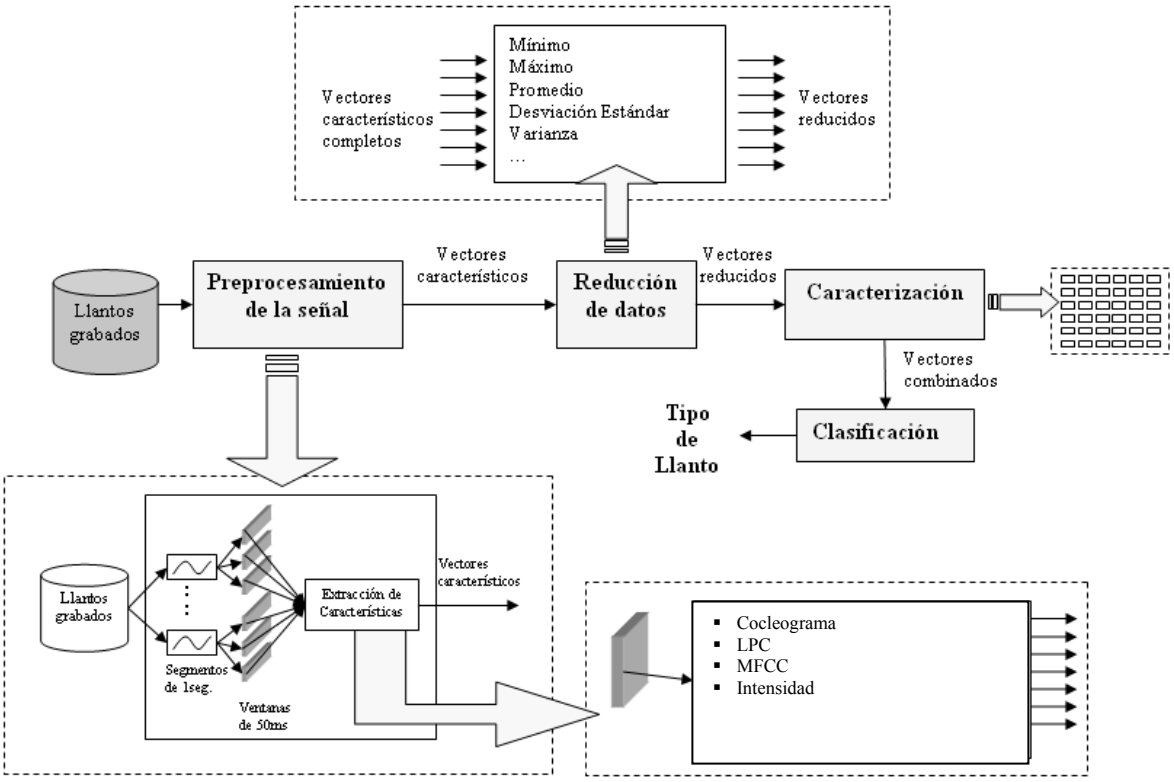


Figura 4.1.1. Modelo global del sistema.

Una descripción general del funcionamiento del sistema propuesto se describe a continuación y más adelante se detallan cada uno de los módulos que lo conforman.

El sistema visto como un modelo global, recibe como entrada las grabaciones de los llantos de bebés, cada grabación es dividida en segmentos de un segundo, posteriormente, cada segmento pasa al módulo de extracción de características donde a través de diversas técnicas se extraen características de la señal y se genera una matriz por cada técnica aplicada, estas matrices de vectores característicos pasan al siguiente módulo que es el de reducción de datos, en el cual las matrices son reducidas, aplicando en este caso dos métodos de reducción (operaciones estadísticas y LDA), los vectores reducidos pasan al siguiente módulo que es el de caracterización, en el cual, la información de los vectores de diferentes características se combinan y generan nuevas matrices de datos con vectores reducidos y combinados. Finalmente, los diferentes tipos de características y sus combinaciones son probados por medio de un clasificador, el cual, al final de un proceso de entrenamiento y prueba, define la precisión en cuanto al reconocimiento para el tipo de llanto que se desea identificar.

A continuación se describe con detalle de cada uno de los módulos.

4.2 Base de llantos

La colección de muestras de llanto de bebés se formó con grabaciones de distinta procedencia. Un primer grupo, (que fue la base para trabajos anteriores) fue la colección de muestras grabadas en el Instituto Nacional de la Comunicación Humana (INCH), en la Ciudad de México y en el Instituto Mexicano del Seguro Social (IMSS) en Puebla, entre ambos se logró una colección de 116 grabaciones correspondientes a 53 bebés (normo-oyentes, con hipoacusia y asfixia). En el INCH se usó un equipo Fisher y un micrófono unidireccional MK2; las muestras del IMSS fueron grabadas

con una grabadora digital Sony ICD-67. Para este trabajo, a ese conjunto se agregaron 73 muestras correspondientes a 59 bebés (normo-oyentes e hipoacúsicos) las cuales, fueron grabadas directamente por los médicos del INCH utilizando el mismo modelo de grabadora digital. Finalmente, un tercer grupo de 6 muestras fue agregado de un conjunto de llantos patológicos, del cual se seleccionaron aquellas grabaciones de bebés con hipoacusia. La Tabla 4.2.1 muestra el conjunto total de muestras de cada clase.

Tabla 4.2.1. Número de grabaciones totales de la base de llantos.

Clase	Grabaciones originales	Detalle
Asfixia	6	
Sordos	66	
Normal	123	Baño 4 Dolor 26 Hambre 36 Incomodidad 8 Normal 26 Otros 23
Total	**195**	

4.3 Procesamiento de la señal

Debido a que las muestras recolectadas fueron grabadas y digitalizadas bajo condiciones diversas y con parámetros distintos, fue necesario remuestrearlas para tener uniformidad, es decir, que la frecuencia y resolución del audio fueran las mismas para todas y cada una de las muestras. El remuestreo se refiere a la operación de modificar la frecuencia de muestreo de un archivo de audio, sin alterar la frecuencia del sonido (mantener el tiempo). Esto se consigue normalmente eliminando o repitiendo algunas muestras. Para pasar por ejemplo, de 44.100 Hz a 22.050, se elimina directamente una muestra de cada dos, mientras que para realizar el cambio inverso, cada muestra es duplicada. En realidad, para obtener una mayor calidad se realiza una interpolación, de forma que si una muestra vale 1000 y la

siguiente 1020, la que se añade tomará el valor 1010 (cuando el cociente de las dos frecuencias no es un valor entero, las matemáticas involucradas se complican un poco más, pero el principio sigue siendo el mismo). En este trabajo, los valores originales de la frecuencia de muestreo que se obtuvieron de cada archivo de audio estuvieron en el rango de [8000-44100] Hz con resoluciones de 8 y 16 bits. Para elegir los parámetros adecuados de la frecuencia de muestreo y la resolución en el proceso de remuestreo, se obtuvieron las propiedades de cada muestra y se opto por mantener 8000Hz y una resolución de 8bits, que fueron la frecuencia mínima y la resolución mínima.

A partir de las muestras con parámetros uniformes se realizó el proceso de segmentación, para el cual se programó un *"script"* en PRAAT que realizara dicho proceso de manera automática, ya que en trabajos anteriores [12] [16] este proceso se llevaba a cabo casi de manera manual.

En el proceso de segmentación, fue importante considerar el tamaño del segmento, ya que de él depende el número de muestras (o segmentos) a procesar y también el número de parámetros a obtener de cada muestra. En el trabajo de Orozco [12], se llevaron a cabo pruebas con segmentos de 1 y 3 segundos, obteniendo los mejores resultados en el proceso de clasificación con muestras de 1 segundo. En [16] de igual manera la segmentación de 1 segundo obtuvo buenos resultados, por lo que en este trabajo se decidió realizar la segmentación de los archivos de audio en muestras de 1 segundo de duración.

Cada grabación fue dividida en segmentos de 1 segundo, tomando la base de llantos formada por 195 grabaciones de diferentes longitudes en tiempo, variando el número de segmentos obtenidos en el rango de 1 a 180, dependiendo del tiempo de duración (en segundos) de cada grabación. Al final del proceso de segmentación se generaron: 4430 muestras de 1 segundo de llanto de bebé normal, 2809 de llanto de bebé sordo y 340 de asfixia, como se muestra en la Tabla 4.3.1.

Tabla 4.3.1. Número de segmentos que se obtendrían con diferentes tamaños de segmento (tiempo).

Tamaño del segmento	Cantidad de segmentos generados por cada clase		
	Normal	Sordos	Asfixia
1 segundo	4430	2809	340

4.3.1 Módulo de extracción de características

La extracción de características se realizó sobre cada muestra de 1 segundo de la siguiente manera: considerando un tamaño de ventana de 50ms, cada muestra de 1 segundo fue subdividida en ventanas, sobre las cuales se aplicaron cuatro técnicas de extracción de características: MFCC, LPC, intensidad y cocleograma, (ver Figura 4.3.1.1). En general, para cada técnica de extracción se utilizaron los valores por defecto del programa PRAAT, haciendo algunas modificaciones para algunos valores de sus parámetros (más adelante se explican las razones de cada modificación). Los tipos de características obtenidas y los valores para los parámetros que se modificaron se muestran en la Tabla 4.3.1.1.

Tabla 4.3.1.1. Técnicas de extracción de características y parámetros modificados.

Técnica de extracción	Parámetros	
MFCC	No. Coeficientes	16
	Tamaño de la ventana	50ms=0.05
	Paso (*Time Step*)	0.05
LPC	Tamaño de la ventana	50ms=0.05
	Paso (*Time Step*)	0.05
Intensidad	Tamaño de la ventana	50ms=0.05
	Paso (*Time Step*)	0.05
Cocleograma	Paso (*Time Step*)	0.1

La extracción de características en trabajos previos se realizaba casi de manera manual, por lo que se programó un *script* en PRAAT que automatizó este proceso.

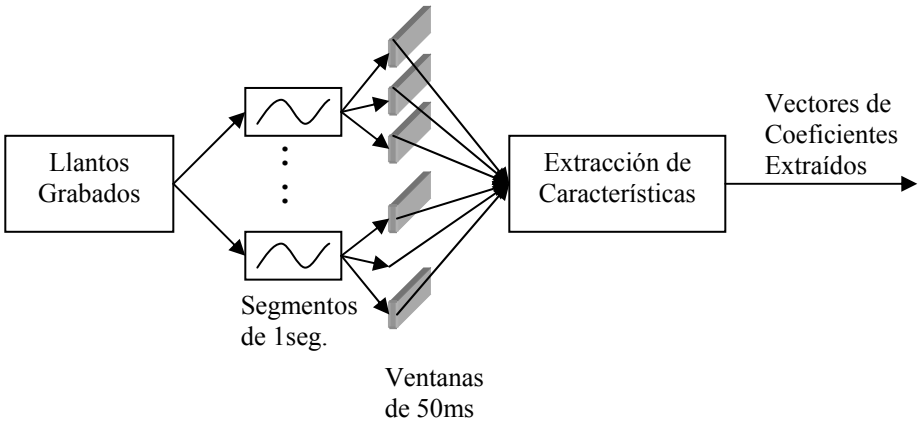


Figura 4.3.1.1. Proceso de extracción de características

El número de parámetros finales obtenidos por cada técnica de extracción aplicada a cada muestra de 1 segundo se obtuvo de la siguiente manera:

- Para MFCC y LPC, con un *Time Step* de 50ms=0.05, 19 es el número de pasos (que obtiene PRAAT) en 1 segundo. Se extraen 16 coeficientes cada paso tomando una ventana de análisis de 50ms, multiplicando 19x16 tenemos un total de 304 coeficientes en total para cada muestra de 1 segundo.

- Para la intensidad, igual que con MFCC y LPC, se definió un *Time Step* de 50ms. Por cada ventana se calcula un valor de intensidad, por lo que finalmente se tuvieron 19 valores de intensidad por cada muestra de 1 segundo.

- Para calcular el número de parámetros a obtener del cocleograma, se divide el valor de frecuencia más alta (25.6 Bark) entre el valor de la resolución (25.6/0.1= 256) lo cual define el número de bandas de frecuencia, y el resultado se multiplica por el número de pasos que resultan de dividir el tiempo (1 segundo) entre el valor de *time step* (1/.1=10), esto es: 256x10=2,560 parámetros de excitación, lo cual es 10 veces menor que

25,600, aunque sigue siendo un valor muy grande de parámetros se optó por mantenerlo de ésta manera.

A continuación se explican las razones para modificar los valores de los parámetros que maneja por defecto PRAAT.

El número de coeficientes MFCC se estableció con un valor de 16, PRAAT maneja por defecto el valor de 12 para trabajar con habla. Sin embargo, dado que en trabajos previos [12] [16] se demostró que con 16 coeficientes se obtienen los mejores resultados en la clasificación del llanto, se decidió no descartar dichas aportaciones y mantener el valor de 16 coeficientes MFCC.

Para el valor del tamaño de la ventana, en reconocimiento de voz, típicamente se maneja un tamaño de 20-30ms. Sin embargo, en llanto de bebé, las variaciones son más lentas por lo que se define un tamaño de ventana un poco más grande, en este caso se ha observado que 50ms, es un valor apropiado para el tamaño de la ventana.

El valor del paso o *time step*, está ligado al valor del tamaño de la ventana, ya que éste, define el paso entre dos ventanas de análisis consecutivas, en otras palabras, el periodo de muestreo. Para los experimentos realizados en este trabajo, a excepción del cocleograma, se definió el valor para *time step* de 0.05= 50ms (el mismo tamaño de la ventana) es decir, no existe traslape entre ventanas. Si se quisiera cierto traslape, el valor del paso debería ser un valor menor al tamaño de la ventana y entonces, probablemente, también habría un incremento en el número de ventanas de análisis que se obtendrían y por lo tanto también un incremento en el número de parámetros extraídos.

Para el caso del cocleograma se definió un *time step* con un valor más grande igual a 0.1=100ms, el valor que manejaba PRAAT era de 0.01=10ms, el cual era muy pequeño y se obtenían 25,600 parámetros (obtenidos con el valor por defecto para la

resolución de frecuencia 0.1 y una ventana de 0.03). Al cambiar el valor de *time step* a 0.1 se reduce el número de parámetros extraídos a 2560 (que son los que se utilizan en este trabajo). Si se quisiera disminuir aun más éste número, se tendría que modificar el valor de la resolución que por defecto es 0.1 a un valor más grande. Por ejemplo, si se definiera la resolución con 0.5, tendríamos 25.6/0.5= 51.2 ≈ 51 bandas de frecuencia en lugar de 256. Calculando 10 parámetros por banda (*time step*= 0.1), se obtendrían 510 atributos.

4.4 Módulo de reducción de datos

Existen diversos métodos para llevar a cabo la reducción de datos, en este caso se analizaron tres de ellos: Análisis Discriminante Lineal (LDA), Análisis de Componentes Principales (PCA) y reducción por operaciones estadísticas. Los dos primeros, son dos métodos ampliamente utilizados en diversos campos y, la reducción por operaciones estadísticas es un método que se propone en este trabajo de tesis. Después de llevar a cabo una serie de pruebas, se seleccionaron dos métodos de reducción: LDA y reducción por operaciones estadísticas, ya que fueron los que obtuvieron los mejores resultados. Ambos métodos se aplicaron sobre los vectores de datos previamente extraídos por las técnicas de extracción de características. A continuación se describe cada uno de ellos.

4.4.1 LDA

Análisis Discriminante Lineal (LDA), es un método comúnmente usado para llevar a cabo clasificación y reducción de datos. Este método maximiza el radio de la varianza entre-las-clases y la varianza dentro-de-las-clases en cualquier conjunto de datos particular de tal modo que garantiza la separación máxima. En el método de análisis de componentes principales (PCA), por ejemplo, la forma y localización de los datos es cambiada cuando se transforman a diferentes espacios [35], LDA no cambia la

localización de los datos, sino trata de encontrar la separabilidad dibujando una región de decisión entre las clases dadas.

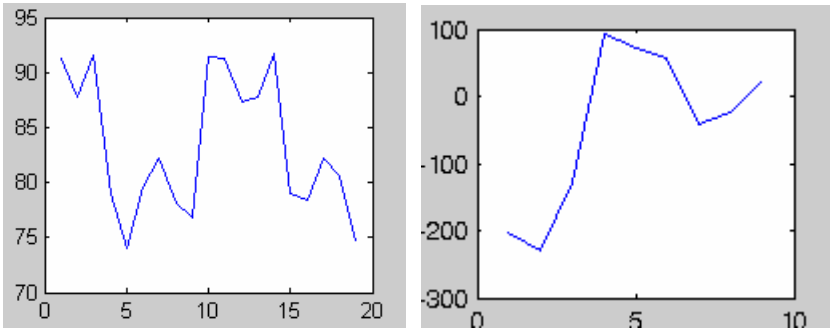
Para seleccionar el número adecuado de LDA's, se aplicó la reducción desde 1 hasta 20 LDA's sobre vectores de características obtenidos por medio de tres técnicas de extracción. Posteriormente, se realizaron pruebas de clasificación (P1, P2, P3) para cada técnica, utilizando validación cruzada, la Tabla 4.4.1.1 muestra los resultados de las pruebas realizadas, ordenadas ascendentemente de acuerdo a su precisión global promedio.

Tabla 4.4.1.1 Pruebas para la selección del número de atributos a utilizar con LDA.

No. LDA's	Intensidad			LPC			MFCC			% Global
	P1	P2	P3	P1	P2	P3	P1	P2	P3	
15	83,66	84,47	86,89	95,96	95,31	93,37	92,72	95,79	93,69	**91,32**
16	84,14	84,63	84,47	93,69	93,85	93,85	95,47	93,85	96,44	**91,15**
20	86,25	85,76	84,79	94,01	92,07	93,37	93,69	95,31	92,72	**90,88**
9	83,33	84,14	84,3	93,85	94,5	94,82	93,04	93,2	94,82	**90,67**
13	82,04	86,41	85,11	93,69	93,53	92,72	94,01	94,34	93,53	**90,60**
18	86,89	83,5	86,41	92,72	93,37	93,53	92,4	92,56	92,72	**90,45**
10	85,44	82,69	83,33	92,56	93,69	93,85	94,18	94,66	93,37	**90,42**
11	82,85	83,5	83,5	92,56	92,56	93,37	94,34	94,34	94,18	**90,13**
14	83,82	84,14	82,69	94,66	93,37	93,37	90,45	94,5	92,56	**89,95**
17	83,5	78,64	83,33	93,69	94,5	94,82	93,37	93,04	94,18	**89,90**
8	80,58	82,69	80,1	94,82	93,37	93,85	93,04	94,5	94,82	**89,75**
12	83,33	81,88	84,47	94,34	91,91	94,5	90,62	90,78	93,04	**89,43**
19	84,95	74,27	84,14	92,4	93,53	92,72	92,4	96,12	93,85	**89,37**
5	79,94	77,83	75,41	95,79	93,69	94,34	95,63	95,31	93,85	**89,09**
3	79,61	78,32	76,7	93,37	94,98	94,34	94,82	94,34	93,69	**88,91**
4	78,8	72,17	77,67	94,34	95,31	95,31	93,53	94,01	94,66	**88,42**
7	80,1	67,31	79,94	93,2	93,53	93,04	94,5	94,34	94,5	**87,83**
1	72,82	74,43	76,21	94,98	95,63	94,34	94,98	93,2	93,04	**87,74**
6	77,02	78,64	66,99	93,37	94,66	94,18	94,01	92,07	94,5	**87,27**
2	70,23	74,92	70,87	93,69	96,76	94,01	92,88	93,69	95,47	**86,95**

Los mejores resultados se obtuvieron usando 15, 16, 20 y 9 LDA's. De los cuales, se decidió utilizar 15 y 9 LDA's, 15 por obtener el mejor resultado y 9 por tener el mejor resultado con el menor número de LDA's, y además se observó que la diferencia era mínima comparando contra los resultados de 16 y 20 LDA's.

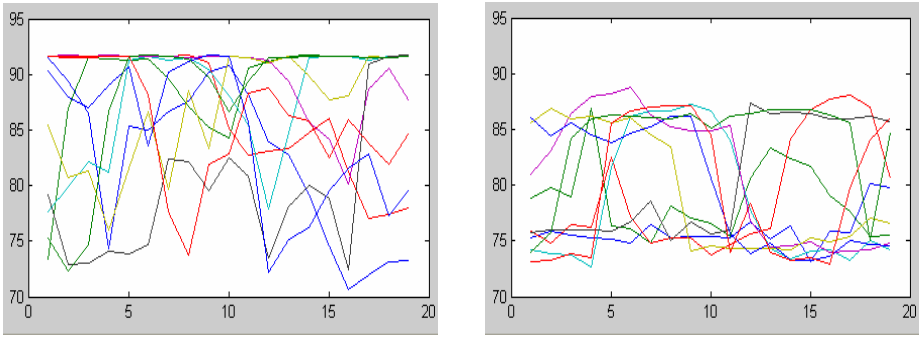
Para ilustrar el resultado de la reducción, a continuación se muestra un ejemplo tomando un vector característico de intensidad, el cual se grafica en la Figura 4.4.1.1. (a), y su representación con 9 LDA's en (b) de la misma figura.



(a) (b)

Figura 4.4.1.1. (a) Gráfica de un vector de intensidad y (b) su reducción por 9 LDA's .

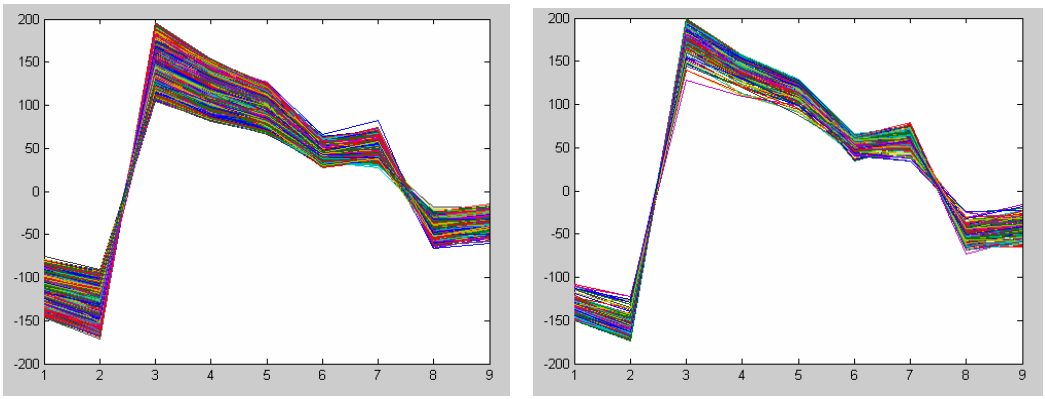
La Figura 4.4.1.2. (a) muestra la representación gráfica de algunos vectores de intensidad de la clase normal, y en (b) de la clase sordos.



(a) (b)

Figura 4.4.1.2 Representación gráfica de vectores de intensidad de la clase normal (a) y sordos (b)

Al observar la Figura 4.4.1.2 es difícil determinar algún patrón que nos ayude a diferenciar una clase de otra, sin embargo, al visualizar la representación de los espacios reducidos, se pueden apreciar diferencias en cuanto al rango de valores que abarcan cada una de las clases. La Figura 4.4.1.3 muestra la representación del espacio reducido para ambas clases, graficando los vectores característicos de intensidad reducidos por LDA de la clase normal en (a) y en (b) para la clase sordos.



(a) (b)

Figura 4.4.1.3 Representación gráfica de vectores de intensidad de la clase normal (a) y sordos (b) reducidos por LDA.

En la Figura 4.4.1.3 se observa que hay diferencias mínimas entre los valores que se obtuvieron para cada clase al aplicar LDA. Sin embargo, esas pequeñas diferencias son las que el clasificador aprovechará para diferenciar entre una clase y otra. Más adelante se muestran los resultados que se obtuvieron en los experimentos de clasificación.

4.4.2 Reducción por operaciones estadísticas

La reducción utilizando operaciones estadísticas es un método propuesto en esta tesis. Es un método relativamente sencillo en comparación con otros como LDA o PCA, los cuales requieren información de todas las clases a reducir para encontrar los

parámetros o componentes que mejor representen cada una de ellas. En cambio, la reducción por operaciones estadísticas se lleva a cabo independientemente de las clases, es decir, se aplica de manera local a cada segmento en particular sin considerar a que clase pertenece. El método consiste en aplicar 5 operaciones estadísticas sobre cada vector de características extraído, las operaciones estadísticas son: mínimo, máximo, promedio, desviación estándar y varianza. Los parámetros estadísticos (al igual que con LDA) se calculan después de que la señal ya fue analizada por alguno de los métodos de extracción de características. Por ejemplo, un vector característico obtenido por LPC, al reducirlo por operaciones estadísticas estaría formado de la siguiente manera: en el parámetro uno, el mínimo de los coeficientes LPC; el segundo parámetro, el máximo de los coeficientes LPC; en el tercer parámetro estadístico estaría el promedio de los valores de los coeficientes; el cuarto parámetro sería la desviación estándar y el quinto la varianza. Recordando que el proceso de extracción de características con LPC, fue sobre segmentos de 1 segundo resultando en 304 coeficientes para cada segmento, con reducción estadística los 304 coeficientes se reducen a sólo 5. La Tabla 4.4.2.1 muestra el número de atributos a reducir para cada vector característico.

Tabla 4.4.2.1 Número de atributos a reducir para cada vector característico.

Tipo de características	No. de atributos a reducir
LPC	**304**
MFCC	**304**
Intensidad	**19**
Cocleograma	**1500**

Para ilustrar el proceso de reducción se presenta el siguiente ejemplo:

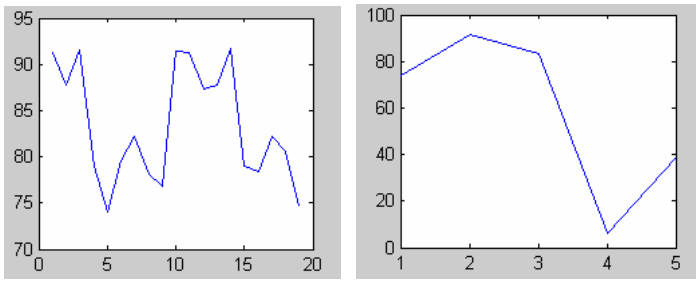
Se tiene el vector de intensidad del llanto de un bebé normal:

91.3	87.7	91.6	78.9	74.1	79.5	82.1	78.2	76.8	91.5	91.3	87.3	87.7	91.7	79.0	78.4	82.2	80.6	74.6	1

El vector estadístico reducido es el siguiente:

Mínimo	Máximo	Promedio	D. Estándar	Varianza	Clase
74.08	91.70	83.40	6.24	39.05	1

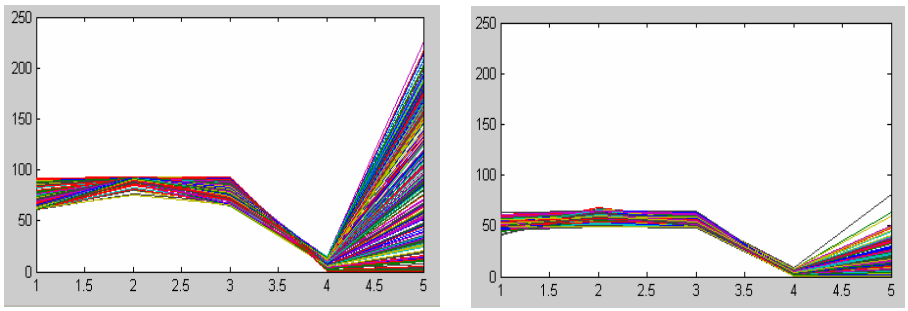
La Figura 4.4.2.1 (a) muestra la representación gráfica de dicho vector y en (b) la representación gráfica de su reducción.



(a) (b)

Figura 4.4.2.1 (a) Gráfica de un vector de intensidad y (b) su reducción por operaciones estadísticas.

La representación gráfica de todo el conjunto de vectores de intensidad reducidos por operaciones estadísticas de la clase normal en (a) y sordos (b) se muestran en la Figura 4.4.2.2.



(a) (b)

Figura 4.4.2.2 Representación gráfica de vectores de intensidad de la clase normal (a) y sordos (b) reducidos por operaciones estadísticas.

Las gráficas para LPC, MFCC y cocleograma, se muestran en el Apéndice A.

4.4.3 Comentarios del proceso de reducción

En las gráficas de los datos sin reducción no se puede determinar si existe algún patrón en cuanto al comportamiento de los datos obtenidos. Sin embargo, al observar las gráficas de las reducciones tanto de LDA como de operaciones estadísticas, se ve claramente que existe un comportamiento bien definido en los parámetros extraídos. Es importante mencionar que el método de reducción estadística a pesar de que puede ser visto como un método "drástico" al reducir los datos, los parámetros que se mantienen al final representan características importantes. En el caso del mínimo y el máximo, son datos que pasan sin sufrir ninguna modificación y lo que definen en su combinación es el rango de valores para todo el vector de datos, en el caso del promedio, la desviación estándar y la varianza, son datos que toman en cuenta a todos y cada uno de los valores del vector, y que a su vez de manera independiente, cada uno aporta información que representa el comportamiento de cómo están distribuidos y separados los datos.

En el caso de la desviación estándar y la varianza, abría que llevar a cabo un análisis para determinar que parámetro aporta más información al clasificador, ya que son datos que están estrechamente correlacionados. Sin embargo, en este trabajo no se consideró dicho análisis y por lo tanto las pruebas se aplicaron considerando ambos valores.

En el caso del cocleograma, principalmente para la clase sordos, al inicio y al final de los vectores existían atributos con valor cero, los cuales no aportaban información de utilidad, sobre todo para la reducción estadística. La solución en este caso fue seleccionar un rango más pequeño (de los 2560 atributos), donde se excluyeran la mayoría de valores con ceros. El rango seleccionado fue del 501 al 2000, esto es, ignorar los primeros 500 valores y los últimos 560. Por tanto, el tamaño de los vectores a reducir para el cocleograma fue de 1500 atributos, de los cuales, existía la posibilidad de que aún hubiera valores con ceros, por lo que para obtener el valor

mínimo, se ordenó el vector y se tomó el primer valor más pequeño diferente de cero[6].

4.5 Combinación de características

Dentro del área del reconocimiento automático del llanto de bebé, la combinación de características en un vector es una idea que surge y se plantea como trabajo futuro en trabajos previos [12] [16] para mejorar la representación de las clases de llanto. En este trabajo, además de realizar pruebas con los diferentes tipos de características de manera individual (pero ahora con una base de grabaciones mayor), se propone llevar a cabo la combinación de diferentes características en un solo vector. Basados en la idea de que cuando las características basadas en predecir la señal, y las basadas en percepción se combinen, se tendrá una representación con atributos que se complementen entre sí, y en consecuencia se mejore la tasa de reconocimiento final.

El número de combinaciones que se puede construir se calcula mediante la fórmula:

$$C_{\mathrm{m}}^{\mathrm{n}} = \binom{m}{n} = \frac{m!}{(m-n)!n!}$$

Para 4 tipos de características, haciendo las combinaciones de dos, tres y cuatro tipos distintos tenemos un total de 11 combinaciones:

$$C_2^4 = \binom{2}{4} = \frac{4!}{(4-2)!2!} = \frac{24}{2*2} = 6$$

$$C_3^4 = \binom{3}{4} = \frac{4!}{(4-3)!3!} = \frac{24}{1*6} = 4$$

$$C_4^4 = \binom{4}{4} = \frac{4!}{(4-4)!4!} = \frac{24}{1*24} = 1$$

[6] Todos los valores obtenidos por el cocleograma son positivos.

Las combinaciones se realizaron con vectores reducidos generando 11 combinaciones por cada método de reducción. La Tabla 4.5.1 muestra la lista de las combinaciones generadas.

Tabla 4.5.1 Matrices generadas mediante combinación de características.

2 tipos de características	3 tipos de características	4 tipos de características
LPC-MFCC	LPC-MFCC-Cocleograma	LPC-MFCC-Cocleograma-Intensidad
LPC-Cocleograma	LPC-MFCC-Intensidad	
LPC-Intensidad	MFCC-Cocleograma-Intensidad	
MFCC-Cocleograma	Cocleograma-Intensidad-LPC	
MFCC-Intensidad		
Cocleograma-Intensidad		

En total se generaron 33 matrices de vectores combinados, 11 por cada método de reducción: estadísticas, 9 LDA's y 15 LDA's.

La forma de llevar a cabo la combinación de características es la siguiente: se toman los *n* atributos reducidos (*n* = 5, 9 ó 15) de cada tipo de característica extraída por las diferentes técnicas y se combinan en un solo vector. Por ejemplo, para generar la combinación LPC-MFCC utilizando parámetros estadísticos, se toman los 5 atributos de LPC, los 5 de MFCC y se unen en un solo vector como se muestra a continuación[7]:

Min LPC	**Max LPC**	**Prom LPC**	**D.Est. LPC**	**Var. LPC**	Min MFCC	Max MFCC	Prom MFCC	D.Est. MFCC	Var. MFCC	Clase

Al final del vector se agrega la clase correspondiente a tipo de llanto al que pertenecen dichos atributos. La Tabla 4.5.2 muestra las dimensiones de las matrices generadas de acuerdo al número de características de diferentes tipos que se

[7]En el caso de los datos reducidos por LDA, en lugar de 5 se toman 9 y 15 atributos.

combinaron y al método de reducción. A la Tabla 4.5.2 además, se agregan las matrices de características sin combinación.

Tabla 4.5.2. Dimensiones de las matrices generadas para cada método de reducción.

Características	Estadística		LDA 9		LDA 15	
	Normal	Sordos	Normal	Sordos	Normal	Sordos
Aplicando 1 tipo de características	5x4430	5x2809	9x4430	9x2809	15x4430	15x2809
Combinación de 2 tipos de características	10x4430	10x2809	18x4430	18x2809	30x4430	30x2809
Combinación de 3 tipos de características	15x4430	15x2809	27x4430	27x2809	45x4430	45x2809
Combinación de 4 tipos de características	20x4430	20x2809	36x4430	36x2809	60x3620	60x2809

4.6 Clasificación

Para la fase de clasificación se eligió como clasificador a la Red Neuronal *Feed Forward Backpropagation,* ya que fue el clasificador que mejores resultados obtuvo comparado contra otros modelos que fueron probados. Los criterios que se tomaron en cuenta para la selección fueron principalmente la precisión y en algunos casos se consideró además el tiempo y la convergencia durante el entrenamiento. A continuación se describe de manera general el proceso de selección del clasificador.

Para probar los distintos modelos, se generaron dos conjuntos, un conjunto *A*, formado por 200 muestras aleatorias de llanto de tipo normal, 200 muestras aleatorias de llanto de tipo sordos y 200 muestras de llanto tipo asfixia. De manera similar, un conjunto *B* se formó con 340 muestras de cada tipo de llanto. Ambos conjuntos utilizaron matrices de datos formadas por vectores característicos MFCC reducidos por operaciones estadísticas.

Entre los clasificadores probados estuvieron: Máquinas de Soporte Vectorial (SMO) [26], Red Neuronal Feed-Forwad Backpropagation, C4.5 [27], Random Forest [28] y Naive Bayes [29]. Además, se probaron algunos ensambles combinando los

clasificadores de mejores resultados bajo diferentes enfoques como, Voto Mayoritario [30], Staking [31], Bagging [30] y Boosting [30]. Otros modelos probados fueron ANFIS [32] y la Red Neuronal Input-Delay [33]. La Tabla 4.6.1 muestra los resultados obtenidos con clasificadores individuales y la Tabla 4.6.2 los obtenidos usando ensambles, en ambos casos, se aplicó validación cruzada de 10 *folds*[8].

Tabla 4.6.1 Clasificadores individuales probados.

Clasificador	Precisión A	Precisión B
Naive Bayes	87.7%	88.4%
SMO	89.7%	90.8%
R.N. Feed-Forwad	**91.7%**	**91.8%**
Random Forest	90.3%	91.4%
J48	89%	90.9%
ANFIS	88.5%	89.5%
R.N. Input-Delay	90.6%	90.8%

Tabla 4.6.2. Ensambles

Ensamble	Precisión A	Precisión B
Staking: Random Forest, SMO	90.8%	92.4%
Staking: SMO, Random Forest	91%	92.1%
Staking: Neural N, SMO, Random Forest	91.7%	92.9%
Staking: Naive Bayes, Random Forest, SMO.	**91.8%**	90.4%
AdaBoost: SMO	89.8%	74.0%
AdaBoost: Random Forest	90.5%	92.5%
AdaBoost: Naive Bayes	88%	88.5%
AdaBoost: J48	90.8%	90.2%
Staking: SMO, J48, Naive Bayes	89.8%	92.0%
Bagging: J48	90.3%	92.5%
Bagging: SMO	89.8%	90.2%
Bagging: Bayes	87.5%	88.4%
Vote: Bayes, SMO, J48	91%	91.1%
Vote: Neural N, Random Forest	91.7%	**93.2%**
AdaBoost: Neural N.	91.7%	91.3%

[8] Los resultados de estas pruebas tanto de los clasificadores individuales como de ensambles fueron publicados en la referencia [36] de esta tesis.

Los ensambles de clasificadores surgen con la idea de mejorar la precisión que se tiene al utilizar un solo clasificador, sin embargo, el tiempo que requieren para entrenar y probar sus modelos es mayor. Además, en las pruebas que se realizaron no se encontraron diferencias significativas comparando el mejor resultado para el conjunto *A* de 91.8% obtenido con el ensamble Staking: NaiveBayes-RandomForest - SMO contra el 91.7% obtenido con la Red Neuronal Feed-Forwad cuya diferencia fue de 0.1%. Por otro lado, para el conjunto *B*, el ensamble Vote: Neural N-Random Forest obtuvo el 93.2%, representando un incremento del 1.4% con respecto al mejor resultado utilizando la Red Neuronal con 91.8%, pero no tuvo ninguna ganancia para el conjunto *A*. La decisión final fue utilizar la Red Neuronal *Feed Forward* ya que para ambos conjuntos de datos fue el clasificador que obtuvo los mejores resultados.

4.6.1 Parámetros de entrenamiento y pruebas de la Red Neuronal

Cualquier proceso de clasificación que utilice redes neuronales consta de dos fases: la fase de aprendizaje o entrenamiento y la fase de prueba. A continuación se describe cada una de ellas.

4.6.1.1 Fase de aprendizaje o entrenamiento

En la fase de entrenamiento se usa un conjunto de datos o patrones de entrenamiento para determinar los pesos que definen el modelo neuronal. Este modelo una vez entrenado, se usa en la fase de prueba.

El número de neuronas de las capas de entrada y salida depende de cada aplicación en particular. Sin embargo, aunque el funcionamiento de la red depende en forma importante del número de nodos en las capas ocultas, no existe aun un método confiable que permita determinar con precisión el número óptimo de estos, aun para alguna o algunas aplicaciones en particular. Una manera de estimar el número óptimo de nodos en la capa oculta es detener el entrenamiento después de un cierto número

de iteraciones y determinar cuántos patrones fueron propiamente reconocidos con el número actual de neuronas usadas en la(s) capa(s) oculta(s). Si el resultado de esta prueba no es satisfactorio se agregarán una o más neuronas en la(s) capa(s) oculta(s) para mejorar el desempeño de la red [38].

La **configuración** de la red neuronal que se utilizó en todas las pruebas fue la siguiente:

- Neuronas en la capa de entrada: 5, 9 y 15.
- Número de capas ocultas: 2.
- Número de neuronas por capa oculta: 10.

Entrenamiento:

- Función de entrenamiento: Levenberg- Marquardt Backpropagation [34].
- Criterios de paro: 50 épocas o convergencia con un error de 0.1, 0.001, 0.0001 (dependiendo de los datos, las combinaciones convergen más rápido por lo que se tuvo que ajustar el valor del error permitido para dar mayor tiempo al entrenamiento).

4.6.1.2 Fase de Prueba

Los pesos de la red neuronal se obtienen a partir de patrones de entrenamiento. Una vez calculados los pesos de la red, se comparan las salidas deseadas con los valores de salida de las neuronas de la última capa para determinar la validez del modelo generado.

Capítulo 5

EXPERIMENTACIÓN Y RESULTADOS

5.1 Descripción de las pruebas realizadas

Inicialmente se realizó la clasificación para dos clases: normal y sordos, con un total de 4430 segmentos o muestras de clase normal y 2809 de clase sordos. Posteriormente, se agregó la clase asfixia de la cual se obtuvieron 340 muestras de 1 segundo. Para llevar a cabo el proceso de clasificación se consideraron diversas pruebas, las cuales se describen a continuación.

5.1.1 Pruebas globales por conjunto de entrenamiento y prueba

Estas pruebas se realizaron para clasificar llanto tipo normal y sordo. Se tomaron aproximadamente el 70% del total de muestras para entrenamiento (seleccionadas aleatoriamente) y el 30% restante para el conjunto de prueba. Para tener un balance en el conjunto de entrenamiento y evitar sesgos, se tomó el mismo número de muestras por cada clase, esto es, 2000 vectores de cada clase (normal y sordo) para entrenamiento y 809 de cada clase para el conjunto de prueba. El proceso de selección del conjunto de entrenamiento y prueba se ejecutó 10 veces, en cada uno de los cuales se obtuvo un resultado para la clasificación. El resultado final fue el promedio de los resultados obtenidos en las 10 pruebas.

5.1.2 Pruebas globales con validación cruzada

La técnica de validación cruzada (10 *fold cross validation*) consiste el dividir el conjunto total de muestras en 10 subconjuntos (*folds*). Una vez generados los subconjuntos, se separa un subconjunto para prueba y se entrena con los restantes 9,

se almacena el resultado para ese subconjunto y luego se regresa al conjunto global, luego se separa el siguiente subconjunto, se vuelve a entrenar con los restantes y se prueba, y así sucesivamente hasta probar los 10 subconjuntos. Al final se realiza un promedio con los resultados obtenidos y se obtiene el porcentaje de clasificación final. En este trabajo, para construir la matriz de entrada en este tipo de pruebas, se tomaron 2809 vectores de la clase normal (elegidos aleatoriamente) y 2809 de la clase sordos. Para las pruebas que incluyeron la clase asfixia, se seleccionaron 340 vectores de cada clase.

5.1.3 Pruebas por individuo

Este tipo de pruebas se consideran las más importantes, ya que su objetivo fue simular el proceso de clasificación como se llevaría a cabo en un ambiente real, estas pruebas se aplicaron de manera particular a las clases normal y sordos. Para la clase normal se clasificaron 123 conjuntos de segmentos, correspondientes a 123 grabaciones de este tipo de llanto, 66 conjuntos de segmentos de llanto de bebés sordos y 6 conjuntos de llanto de asfixia, de aquí en adelante a cada conjunto de muestras pertenecientes a una grabación de llanto se le denomina “individuo”. Cada individuo a probar fue separado del conjunto de entrenamiento (de cada clase se tomó un individuo) y posteriormente probado en 5 entrenamientos distintos (o inicializaciones de la red neuronal). Al final de las pruebas, se realizó una votación donde se consideró correcta la clasificación de un individuo si en al menos 4 de las pruebas clasificó correctamente a más del 50% de los segmentos que corresponden al individuo que se probó. En cada una de las pruebas, la matriz de entrenamiento se formó con el tamaño total de la matriz normal – (menos) el número de segmentos correspondientes al individuo a probar, + (más) el tamaño total de la matriz sordos – (menos) el número de segmentos correspondientes al individuo a probar. El conjunto de prueba estuvo constituido por los individuos separados de cada clase.

5.2 Resultados y análisis de los resultados

Se probaron cuatro técnicas de extracción de características (LPC, MFCC, intensidad y cocleograma) de manera independiente así como sus combinaciones (11 combinaciones), esto es: 15 matrices distintas por cada método de reducción que es igual a 45 matrices (15 estadísticas, 15 con 9 LDA's y 15 con 15 LDA's). Para las pruebas globales se utilizaron dos métodos de evaluación (validación cruzada y por conjunto de entrenamiento y prueba) haciendo un total de 90 pruebas globales, considerando dos clases "normal y sordo". Para las pruebas por individuo se consideraron sólo 30 matrices: las reducidas por operaciones estadísticas y las formadas con 9 LDA's.

El propósito de llevar a cabo tres tipos de pruebas, fue abordar (en medida de lo posible) los diversos escenarios que se presentan para validar el modelo propuesto, y de esta manera encontrar las dificultades en cada uno de los casos.

5.2.1 Resultados de las pruebas globales

A continuación se presentan los resultados de las pruebas globales descritas en la sección 5.1.1. y 5.1.2, para clasificar dos tipos de llanto: normal y sordo; comenzando por las pruebas utilizando un tipo de características y posteriormente sus combinaciones.

La Tabla 5.2.1.1 muestra los resultados para las pruebas globales utilizando características distintas para clasificar las clases "normal-sordos". El mejor resultado aplicando reducción por operaciones estadísticas se obtuvo para LPC con un 93.59% de precisión correcta en las pruebas con un conjunto de entrenamiento de 4000 vectores (2000 de cada clase) y 1618 para pruebas (809 de cada clase). Con validación cruzada, nuevamente con LPC se obtuvo el 94.37% de precisión. En el

caso de LDA con 9 atributos, las características que obtuvieron la mejor precisión fueron las del cocleograma con un 94.52% y 95.69% respectivamente para los dos tipos de pruebas. Y finalmente, para 15 atributos LDA en las pruebas sobre un conjunto de entrenamiento las características del cocleograma obtuvieron un 94.56% y con validación cruzada LPC alcanzó el 93.93%.

Tabla 5.2.1.1. Resultados globales para clasificar las clases "normal-sordos" utilizando cuatro tipos de características distintas.

Pruebas con un conjunto de 2000 muestras para entrenamiento y 809 para prueba.
Pruebas aplicando validación cruzada de 10 *folds*.

	ESTADÍSTICA: 5 atributos		**LDA: 9 atributos**		**LDA: 15 atributos**	
LPC	**93.59%**	**94.37%**	94.12%	94.15%	94.22	**93.93%**
MFCC	92.36%	93.12%	93.75%	94.71%	94.06	93.03%
Intensidad	92.72%	92.93%	83.39%	84.46%	84.73	82.32%
Cocleograma	93.47%	94.02%	**94.52%**	**95.69%**	**94.56%**	93.48%

A continuación se presentan los resultados globales utilizando la combinación de dos tipos de características y posteriormente para las combinaciones de tres y cuatro.

En la Tabla 5.2.1.2 se presentan los resultados de la combinación utilizando dos tipos de características, donde se observa un aumento de más de 3 puntos porcentuales usando atributos estadísticos, para ambos tipos de pruebas, y más o menos 2 puntos con LDA. Casi todas las combinaciones resultaron en un aumento en la precisión de la clasificación, sin embargo, la que mayor porcentaje obtuvo fue la combinación de MFCC-Intensidad con un 97.83% aplicando reducción estadística.

Tabla 5.2.1.2. Resultados globales utilizando combinación de dos tipos de características para clasificar las clases "normal-sordos".

Pruebas con un conjunto de 2000 muestras para entrenamiento y 809 para prueba.
Pruebas aplicando validación cruzada de 10 *folds*.

Combinación 2 tipos de características	ESTADÍSTICA: 10 atributos		LDA: 18 atributos		LDA: 30 atributos	
LPC-MFCC	95.52%	95.62%	**96.41%**	**96.61%**	95.57%	**96.42%**
LPC-Intensidad	92.87%	94.64%	95.32%	96.34%	95.46%	96.30%
LPC-Cocleograma	95.06%	94.84%	95.73%	96.16%	**95.75%**	95.87%
MFCC-Intensidad	**97.83%**	96.52%	94.19%	94.73%	93.57%	93.57%
MFCC-Cocleograma	97.28%	**97.58%**	95.40%	96.43%	95.28%	94.82%
Cocleograma-Intensidad	94.16%	96.61%	94.43%	95.09%	92.89%	94.45%

Al combinar tres tipos de características, (Tabla 5.2.1.3.) los parámetros estadísticos volvieron a mostrar un aumento (no tan grande como al pasar de una a dos), y en los parámetros por LDA la diferencia fue mínima, mostrando una disminución en algunos casos. Un dato interesante, es que la combinación que obtuvo el resultado más alto en ambos métodos de reducción fue la generada por las características que obtuvieron el mejor resultado individual combinadas con el mejor resultado de la combinación de dos tipos de características.

Tabla 5.2.1.3. Resultados globales utilizando combinación de tres tipos de características para clasificar las clases "normal-sordos".

Pruebas con un conjunto de 2000 muestras para entrenamiento y 809 para prueba.
Pruebas aplicando validación cruzada de 10 *folds*.

Combinación 3 tipos de características	ESTADÍSTICA: 15 atributos		LDA: 27 atributos		LDA: 45 atributos	
LPC-MFCC-Cocleograma	97.65%	97.78%	**96.39%**	**96.69%**	**96.20%**	**95.89%**
LPC-MFCC-Intensidad	**98.22%**	**98.30%**	95.87%	96.34%	95.67%	95.08%
MFCC-Cocleograma-Intensidad	97.89%	98.03%	95.30%	94.20%	93.61%	95.16%
LPC-Cocleograma-Intensidad	97.79%	97.58%	96.04%	95.89%	95.79%	95.62%

Los resultados de las pruebas con la combinación de cuatro tipos de características se muestran en la Tabla 5.2.1.4 donde se observa un aumento en la precisión (nuevamente con los atributos estadísticos), lo cual nos hace considerar que el cocleograma aportó información útil para diferenciar entre ambos tipos de llanto.

Tabla 5.2.1.4. Resultados globales utilizando combinación de cuatro tipos de características para clasificar las clases "normal-sordos".

▩ Pruebas con un conjunto de 2000 muestras para entrenamiento y 809 para prueba.
□ Pruebas aplicando validación cruzada de 10 *folds*.

Combinación 4 características	**ESTADISTICA: 20 atributos**		**LDA: 36 atributos**		**LDA: 60 atributos**	
LPC-MFCC-Cocle-Intensidad	**98.48%**	**98.66%**	96.15%	95.45%	95.73	96.52%

5.2.2 Resultados de las pruebas por individuo

Para ilustrar este proceso de clasificación, en la Tabla 5.2.2.1 se muestran algunos resultados obtenidos en este tipo de experimentos. Para cada individuo se realizaron 5 pruebas (p1, p2,…, p5), en las cuales, se clasificaron los segmentos de cada individuo. Una prueba suma un punto si se clasifica correctamente a más del 50% de los segmentos que conforman al individuo, por ejemplo, el individuo No.17 de la tabla formado por 5 segmentos, en la prueba 4, clasificó bien sólo 3 segmentos, este valor representa más del 50% del total de segmentos para ese individuo, por lo que se suma un punto a su calificación final que fue de 5 y se considera clasificado correctamente. Contrario al individuo No. 55 que tuvo dos pruebas (p1 y p2) con el 50% o menos de segmentos bien clasificados, lo cual no sumo puntos a su calificación final que fue de 3, con lo cual dicho individuo fue marcado como un error, ya que en este trabajo para que un individuo se considerara correctamente clasificado al menos debería tener una calificación de 4.

Tabla 5.2.2.1. Ejemplo de algunos resultados obtenidos para clasificar el llanto de bebés como individuos para la clase normal, utilizando la combinación de dos tipos de características: MFCC-Intensidad.

No. Individuo	No. Segmentos	p1	p2	p3	p4	p5	%de precisión por muestra	Calificación
1	**17**	17	17	17	17	16	98.824	5
2	**22**	21	20	20	21	21	93.636	5
3	**127**	127	127	127	127	127	100	5
…								
17	**5**	5	5	5	3	5	92	5
18	**89**	87	86	88	86	86	97.303	5
…								
32	**26**	21	22	21	14	23	77.692	5
35	**46**	46	46	46	46	46	100	5
…								
53	**76**	76	75	75	75	76	99.211	5
54	**81**	74	75	72	79	73	92.099	5
55	**8**	4	3	5	5	6	57.5	3
56	**10**	10	10	10	10	10	100	5
…								

El *porcentaje de precisión por muestra* se calcula como referencia para saber con que precisión fue clasificado cada individuo. La Tabla 5.2.2.1 muestra algunos ejemplos. Dicho valor se calcula como el promedio de los resultados de cada una de las pruebas, por ejemplo, para el individuo No. 1 que está constituido por 17 segmentos, en las pruebas 1, 2, 3 y 4 clasificó correctamente el 100% de sus segmentos, sin embargo, en la prueba 5 clasificó correctamente sólo 16 segmentos, lo cual representa el 94.12% de precisión (en esa prueba) con respecto al número de segmentos de esa muestra, es decir, (17-1)x100/17=94.12, calculando el promedio de los resultados para esa muestra en particular tenemos:

$$(100+100+100+100+94.12)/5= 98.82\% \text{ de precisión}$$

Lo cual nos indica que el individuo No. 1 fue clasificado correctamente con una precisión del 98.82% (con respecto al número de segmentos de ese mismo individuo). En el caso del individuo No. 8 se observa que la precisión que obtuvo fue del 57.5%, el cual es un resultado muy bajo, que refleja el por qué el individuo fue marcado

como un error. Se dice que son valores de referencia ya que el porcentaje de precisión que se obtiene, depende del número de segmentos que pertenezcan a cada individuo, ya que, no es lo mismo tener 2 errores de 100, que 2 errores de 10, sin embargo, dichos resultados podrían estandarizarse si en un futuro el número de segmentos por muestra fuera el mismo para todas.

A continuación se muestran los resultados obtenidos en las pruebas por individuo (explicadas en la sección 5.1.3) para cada método de reducción. En general, para calcular el porcentaje de clasificación correcta de cada clase, así como el porcentaje de precisión final global, cada error o individuo mal clasificado se restó como unidad tomando como base el total de llantos de cada clase, es decir, 66 bebés sordos y 123 normales, por ejemplo, en el caso de MFCC (Tabla 5.2.2.1.1.) tuvo 6 errores para la clase normal y 1 error para la clase sordos, entonces se calcula:

(123-errores_normal)x100/123

(66-errores_sordos)x100/66

Sustituyendo tenemos:

(123-6)x100/123= 95.12% de precisión para la clase normal y
(66-1)x100/66= 98.48% de precisión para la clase sordos.

El porcentaje global se obtiene calculando: el total de muestras (de ambas clases), menos la suma de los errores (de ambas), por 100 y el resultado dividido entre el total de muestras, para el mismo ejemplo tenemos:

[(123+66)-7]x100/189=96.3%

5.2.2.1 Clasificación por individuo aplicando reducción estadística

En la Tabla 5.2.2.1.1 se presentan los resultados obtenidos en las pruebas por individuo considerando los diferentes tipos de características de manera individual y reducción por operaciones estadísticas.

Tabla 5.2.2.1.1. Resultados por individuo para clasificar las clases "normal-sordos" utilizando características de diferentes tipos (sin combinación) y reducción por estadísticas.

Tipo de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC	10	91.87	0	**100**	94.71
MFCC	6	95.12	1	98.48	96.30
Cocleograma	**5**	**95.93**	1	98.48	**96.83**
Intensidad	22	82.11	0	**100**	88.36

En la Tabla 5.2.2.1.1 se observa que la mayoría de los errores caen en la clasificación de la clase normal, a pesar de que el número de muestras para entrenar con dicha clase fue siempre superior, pues suponiendo que para el individuo se tuvieran 180 segmentos (el máximo número) quedarían 4250 para entrenar, contra las que quedaran de quitar un individuo a 2809 vectores de la clase sordos. A pesar de ello, vemos que la clasificación de los individuos de la clase sordos, fue en algunos casos perfecta al 100% (LPC e intensidad), y que sólo hubo un error para algunos tipos de características (MFCC y cocleograma). Sin embargo, ninguna de las características que logró el 100% obtuvo la precisión global más alta, que fue con el cocleograma con un 96.83%, de hecho, los dos tipos de características con la clasificación más alta para la clase sordos, fueron las más bajas en la clasificación global. Este tipo de pruebas, a pesar de requerir una gran cantidad de tiempo para probar a cada uno de los individuos, al final lo que permiten ver es realmente donde están reflejados los errores y que características clasifican mejor a una u otra clase. Además, permite identificar las muestras que requieren de un mayor análisis y trabajo para clasificarlas, de las que siempre se clasifican correctamente. Surge entonces la pregunta: si un tipo de características clasifica perfectamente a los sordos y otro tiene

el menor número de errores en la clasificación de la clase normal, entonces, al combinar las características de ambos tipos ¿Se obtendrán mejores resultados globales?, la Tabla 5.2.2.1.2., muestra los resultados de las pruebas combinando dos tipos de características y reducción por operaciones estadísticas.

Tabla 5.2.2.1.2. Resultados por individuo para clasificar las clases "normal-sordos" utilizando la combinación de dos tipos de características y reducción por estadísticas.

.

Combinación 2 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC	7	94.31	0	**100**	96.30
LPC-Intensidad	8	93.50	0	**100**	95.77
LPC-Cocleograma	**4**	**96.75**	0	**100**	**97.88**
MFCC-Intensidad	6	95.12	0	**100**	96.83
MFCC-Cocleograma	5	95.93	0	**100**	97.35
Cocleograma-Intensidad	7	94.31	0	**100**	96.30

Al combinar dos tipos de características, se observa que en todos los casos al comparar los resultados de la combinación con los obtenidos en las pruebas con un solo tipo de características, la disminución del error para clasificar la clase normal se reduce para alguno de los tipos. En la clasificación con un tipo de características, el número de errores más alto era de 22 se redujo a 8, y el error más pequeño que era de 5 errores, disminuyó a 4, aumentando en todos los casos el porcentaje de clasificación global. Además, la combinación de dos tipos de características permitió reducir a 0 los errores en la clase sordos, lo cual representa una aportación significativa para esta investigación. La mejor combinación resultó de combinar las características de LPC con las del cocleograma, la cual obtuvo un 97.88% de individuos clasificados correctamente.

En la combinación de tres tipos de características reducidas por operaciones estadísticas, Tabla 5.2.2.1.3 se observa que se disminuye a 3 el número de errores para la clase normal, y que en el peor de los casos, el número máximo de errores es el

más pequeño de la combinación con 2 tipos de características. Con 0 errores para la clase sordos, se logra un 98.41% de clasificación correcta.

Tabla 5.2.2.1.3. Resultados por individuo para clasificar las clases "normal-sordos" utilizando combinación de tres tipos de características y reducción por estadísticas.

Combinación 3 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC-Cocleograma	3	**97.56**	0	**100**	**98.41**
LPC-MFCC-Intensidad	3	**97.56**	0	**100**	**98.41**
MFCC-Cocleograma-Intensidad	4	96.75	0	**100**	97.88
LPC-Cocleograma-Intensidad	4	96.75	0	**100**	97.88

Finalmente, el resultado de combinar 4 tipos de características se muestra en la Tabla 5.2.2.1.4. En el cual, aplicando reducción estadística, se reduce aun más el error a 2 individuos mal clasificados, logrando un 98.94% global de clasificación correcta.

Tabla 5.2.2.1.4. Resultados por individuo para clasificar las clases "normal-sordos" utilizando combinación de cuatro tipos de características y reducción por estadísticas.

Combinación 4 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC-Cocle-Intensidad	2	**98.37**	0	**100**	**98.94**

5.2.2.2 Clasificación por individuo aplicando reducción por LDA.

Con el fin de tener un punto de comparación para el método de reducción propuesto, se implementó la reducción por Análisis Discriminante Lineal (LDA), el cual es un método de reducción de datos muy utilizado en diversas áreas. En las pruebas globales (validación cruzada y conjunto de entrenamiento-prueba) se observó que la reducción con 15 LDA's no tuvo ventajas comparando los resultados obtenidos con 9 LDA's, por lo que en las pruebas por individuo, sólo se consideró la reducción con 9 LDA's. A continuación se presentan los resultados por individuo aplicando este método de reducción.

Tabla 5.2.2.2.1. Resultados por individuo para clasificar las clases "normal-sordos" utilizando características individuales sin combinación y reducción por 9 LDA's.

Tipo de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC	10	91.87	0	**100.00**	94.71
MFCC	5	**95.93**	1	98.48	**96.83**
Cocleograma	15	87.80	0	**100.00**	92.06
Intensidad	41	66.67	2	96.97	77.25

En la Tabla 5.2.2.2.1 se muestran los resultados por individuo utilizando 9 LDA's. El tipo de características que obtuvo el mejor resultado fue MFCC, con un 96.83% de clasificación correcta. De igual manera que con el método de reducción estadística, gran parte de los errores estuvieron en la clasificación de la clase normal. Lo cual provocó que el porcentaje de clasificación más alto se diera con la característica que tuvo menos errores para clasificar esta clase.

En la combinación con 2 tipos de características, (Tabla 5.2.2.2.2) se observó que nuevamente al combinar el tipo de características que logró el menor número de errores para la clase normal con alguno de los tipos que obtuvieron el menor número de errores para la clase sordos, se obtuvo el resultado más alto, que en este caso fue de 98.94% con la combinación LPC-MFCC.

Tabla 5.2.2.2.2. Resultados por individuo para clasificar las clases "normal-sordos" utilizando combinación de dos tipos de características y reducción por 9 LDA's.

Combinación 2 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC	2	**98.37**	0	**100**	**98.94**
LPC-Intensidad	9	92.68	0	**100**	95.24
LPC-Cocleograma	10	91.87	0	**100**	94.71
MFCC-Intensidad	8	93.50	1	98.48	95.24
MFCC-Cocleograma	6	95.12	0	**100**	96.83
Cocleograma-Intensidad	18	85.37	0	**100**	90.48

Con la combinación de 3 tipos de características y aplicando reducción por 9 LDA's (Tabla 5.2.2.2.3.), se logró el resultado más alto obtenido en todas las pruebas realizadas, reduciendo el error a 1 individuo mal clasificado de la clase normal y 0 de la clase sordos, se obtuvo un 99.47% de precisión.

Tabla 5.2.2.2.3. Resultados por individuo para clasificar las clases "normal-sordos" utilizando combinación de tres tipos de características y reducción por 9 LDA's.

Combinación 3 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC-Cocleograma	1	**99.19**	0	**100**	**99.47**
LPC-MFCC-Intensidad	8	93.50	1	98.48	95.24
MFCC-Cocleograma-Intensidad	4	96.75	0	**100**	97.88
LPC-Cocleograma-Intensidad	6	95.12	0	**100**	96.83

Un dato interesante es que la combinación LPC-MFCC-Cocleograma con reducción estadística también obtuvo el resultado más alto en la combinación de 3 características con un 98.41% de precisión.

Finamente, al agregar la intensidad para formar la combinación de vectores con 4 tipos de características, Tabla 5.2.2.4, la precisión disminuye debido a que las características de intensidad son las que tienen el mayor número de errores para clasificar la clase normal, y también tuvo dos errores en la clasificación de la clase sordos. Por lo cual se deduce que dicho tipo de características, al combinarlas, logran una mejora tomando como referencia las características en sí, pues hay una reducción del error máximo que tenían de 41 individuos de la clase normal mal clasificados a sólo 2. Sin embargo, este tipo de características, en general con LDA, no tuvieron mucha ganancia o beneficio en la combinación.

Tabla 5.2.2.2.4. Resultados por individuo para clasificar las clases "normal-sordos" utilizando combinación de cuatro tipos de características y reducción por 9 LDA's.

Combinación 4 tipos de características	No. individuos mal clasificados clase normal	% clase Normal	No. individuos mal clasificados clase sordos	% clase Sordos	%Global
LPC-MFCC-Cocle-Intensidad	2	98.37	1	98.48	98.41

5.2.3 Pruebas con asfixia

Con el fin de validar el modelo propuesto, se agrega una tercera clase, asfixia. Las pruebas se aplicaron sobre un conjunto de 340 muestras de llanto de cada clase, donde las muestras de las clases sordos y normales fueron elegidas de manera aleatoria de un total de 3340 para la clase normal y 2809 de la clase sordos.

En la Tabla 5.2.3.1 se presentan resultados globales (validación cruzada) para clasificar llanto de asfixia contra llanto normal, utilizando dos métodos de reducción. Se observa que el tipo de características que obtuvieron la precisión más alta fueron las de LPC con un 91.87% con reducción estadística y 91.87% para LDA.

Tabla 5.2.3.1. Resultados para clasificar las clases "asfixia-normal" utilizando cuatro tipos de características y dos métodos de reducción.

Características	Clases	Reducción estadística 5 atributos	Reducción LDA 9 atributos
LPC	Asfixia-normal	**91.87%**	**91.87%**
MFCC	Asfixia-normal	91.5%	89.5%
Intensidad	Asfixia-normal	91%	86.25%
Cocleograma	Asfixia-normal	87.50%	84.37

La Tabla 5.2.3.2 muestra los resultados para clasificar llanto de asfixia y llanto de bebé sordo, obteniendo el mejor resultado las características del cocleograma utilizando reducción estadística logrando un 88.75%, y las características de intensidad aplicando LDA con un 70% de clasificación correcta.

Tabla 5.2.3.2. Resultados para clasificar las clases "asfixia-sordos" utilizando cuatro tipos de características y dos métodos de reducción.

Características	Clases	Reducción estadística 5 atributos	Reducción LDA 9 atributos
LPC	Asfixia-sordos	77.5%	68.75%
MFCC	Asfixia-sordos	87.5%	60%
Intensidad	Asfixia-sordos	81.87%	**70%**
Cocleograma	Asfixia-sordos	**88.75%**	60%

Al clasificar tres clases "normal-sordos-asfixia" se obtuvieron los resultados de la Tabla 5.2.3.3 en la cual el cocleograma logró el 90% de clasificación correcta con reducción estadística, mientras que las características de LPC obtuvieron el 75.83% con reducción LDA.

Tabla 5.2.3.3. Resultados para clasificar tres clases "normal-sordos-asfixia" utilizando cuatro tipos de características y dos métodos de reducción.

Características	Clases	Reducción estadística 5 atributos	Reducción LDA 9 atributos
LPC	Normal-Sordos-Asfixia	82.5%	**75.83%**
MFCC	Normal-Sordos-Asfixia	80%	75%
Intensidad	Normal-Sordos-Asfixia	82.08%	70.83%
Cocleograma	Normal-Sordos-Asfixia	**90%**	72.5%

En la Tabla 5.2.3.4 se presentan los resultados para clasificar 3 clases de llanto "normal-sordo-asfixia" utilizando la combinación de dos, tres y cuatro tipos de características. El mejor resultado para la combinación de dos tipos de características se obtuvo con la combinación MFCC-Cocleograma, para ambos métodos de reducción, con 92% para la reducción estadística y 84.69% con LDA. En la combinación de tres tipos de características, hubo un pequeño aumento, tan solo de unas cuantas décimas, y con cuatro características no se logró mejorar más el resultado obtenido con la combinación de tres características de 92.8% y 84.83% para la reducción estadística y LDA respectivamente.

Tabla 5.2.3.4. Resultados para clasificar tres tipos de llanto "normal-sordos-asfixia" utilizando la combinación de dos, tres y cuatro tipos de características, aplicando dos métodos de reducción.

Combinación 2 tipos de características	**ESTADÍSTICA: 10 atributos**	**LDA: 18 atributos**
LPC-MFCC	87%	81.33%
LPC-Intensidad	89.5%	79.54%
LPC-Cocleograma	87.41%	81.23%
MFCC-Intensidad	87.66%	77.86%
MFCC-Cocleograma	**92%**	**84.69%**
Cocleograma-Intensidad	89.16%	79%
Combinación 3 tipos de características	**ESTADÍSTICA: 15 atributos**	**LDA: 27 atributos**
LPC-MFCC-Cocleograma	**92.8%**	**84.83%**
LPC-MFCC-Intensidad	90.25%	81.9%
MFCC-Cocleograma-Intensidad	91.58%	80.14%
LPC-Cocleograma-Intensidad	91.16%	82.86%
Combinación 4 tipos de características	**ESTADÍSTICA: 20 atributos**	**LDA: 36 atributos**
LPC-MFCC-Cocleograma-Intensidad	91.3%	84.5%

5.2.3.1 Comentarios de las pruebas con asfixia

Un aspecto importante a considerar al agregar una nueva clase al proceso de reconocimiento, es la fase de reducción de datos, pues una vez generados los vectores característicos, la reducción y combinación de características es inmediata aplicando la reducción por estadísticas, ya que únicamente se procesan los nuevos vectores característicos, que en este caso fueron los de asfixia, manteniendo los datos del proceso anterior para normales y sordos sin alterar. Sin embargo para LDA, se requirió de un proceso completo, casi desde cero, pues al incluir nueva información se tuvo que reproducir nuevamente todo el proceso de reducción y combinación de todas las clases involucradas generando nuevas matrices individuales y sus combinaciones para todas las clases.

5.3 Generación de la base de conocimiento

El proceso de construir una base de conocimiento es una parte primordial de la ingeniería de conocimiento, la cual, busca hacer el mejor uso del conocimiento disponible. Su objetivo es capturar, organizar y almacenar la información y experiencias, con el fin de que éstas puedan ser aprovechadas y estén disponibles para otros. Una manera tradicional de representar el conocimiento son las reglas. Una regla es una estructura condicional que relaciona lógicamente la información contenida en la parte del antecedente con otra información contenida en la parte del consecuente, que puede invocarse de manera coordinada para obtener conclusiones o soluciones cuando se plantea una pregunta o un problema. Las reglas representan el conocimiento utilizando un formato SI-ENTONCES (IF-THEN), es decir, tienen 2 partes:

- La parte SI (IF), es el antecedente, premisa, condición o situación; y
- La parte ENTONCES (THEN), es el consecuente, conclusión, acción o respuesta.

Uno de los propósitos de tener una base de conocimiento para clasificar llanto de bebé es que ésta pueda crecer de manera permanente, sin que ello afecte al correcto funcionamiento de la misma y sin necesidad de hacer modificaciones sustanciales en ella. Basados en los resultados obtenidos por los experimentos realizados en este trabajo y trabajos anteriores, para facilitar la comprensión de las reglas por generar, en el Apéndice B se presentan las tablas que resumen los mejores resultados obtenidos por las diversas pruebas antes presentadas. Se utilizará esta información para definir las reglas de la base de conocimiento para clasificar llanto de bebé.

De las Tablas 1B, 2B, 3B y 4B que se muestran en el Apéndice B se deducen las siguientes reglas: (en todos los casos aplicando **reducción estadística**).[9]

- **Si** se desea obtener la mayor precisión para clasificar la **clase "normal"** con **un tipo de características, entonces** se debe utilizar el **Cocleograma o MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar la **clase "sordos"** con **un tipo de características, entonces** se debe utilizar **LPC o Intensidad.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con **un tipo de características, entonces** se debe utilizar **LPC o Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-asfixia"** con **un tipo de características, entonces** se debe utilizar **MFCC o LPC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "sordos-asfixia"** con **un tipo de características, entonces** se debe utilizar **Cocleograma o MFCC.**

[9] La línea continua indica que la misma característica o combinación obtuvo el mejor resultado en ambos métodos de reducción. Para algunos de los resultados, resulto difícil la decisión por una u otra característica o combinación por lo que se decidió mantener la posibilidad para aquellas que obtuvieron los mejores resultados, en este sentido la línea punteada indica que al menos una de las combinaciones o características obtuvo buenos resultados en ambos métodos de reducción. Las reglas sin línea indican que sólo para ese método de reducción en particular obtuvo el mejor resultado.

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordo-asfixia"** con **un tipo de características, entonces** se debe utilizar el **Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con una **combinación de dos tipos de características, entonces** se debe utilizar **cualquier combinación de Cocleograma con LPC o MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos-asfixia"** con una **combinación de dos tipos de características, entonces** se debe utilizar **la combinación de MFCC con Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con una **combinación de tres tipos de características, entonces** se debe utilizar **cualquier combinación de MFCC-LPC con Cocleograma o Intensidad.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordo-asfixia"** con una **combinación de tres tipos de características, entonces** se debe utilizar **LPC-MFCC-Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con una **combinación de cuatro tipos de características, entonces** se debe utilizar la **combinación de MFCC-LPC-Cocleograma-Intensidad.**

De las Tablas 1B, 2B, 3B y 4B que se muestran en el Apéndice B se deducen las siguientes reglas: (en todos los casos aplicando **reducción con LDA**)

- **Si** se desea obtener la mayor precisión para clasificar la **clase "normal"** con **un tipo de características, entonces** se debe utilizar **MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar la **clase "sordos"** con **un tipo de características, entonces** se debe utilizar **LPC o Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con **un tipo de características, entonces** se debe utilizar **Cocleograma o MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-asfixia"** con **un tipo de características, entonces** se debe utilizar **LPC o MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "sordos-asfixia"** con **un tipo de características, entonces** se debe utilizar **Intensidad o LPC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordo-asfixia"** con **un tipo de características, entonces** se debe utilizar **LPC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con una **combinación de dos tipos de características, entonces** se debe utilizar la **combinación de LPC con MFCC.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos-asfixia"** con una **combinación de dos tipos de características, entonces** se debe utilizar **la combinación de MFCC con Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordos"** con una **combinación de tres tipos de características, entonces** se debe utilizar **la combinación LPC-MFCC-Cocleograma.**

- **Si** se desea obtener la mayor precisión para clasificar las **clases "normal-sordo-asfixia"** con una **combinación de tres tipos de características, entonces** se debe utilizar **LPC-MFCC-Cocleograma.**

La representación gráfica de las reglas definidas para ambos métodos de reducción se muestra en las figuras 5.3.1 y 5.3.2 para reducción estadística y LDA respectivamente.

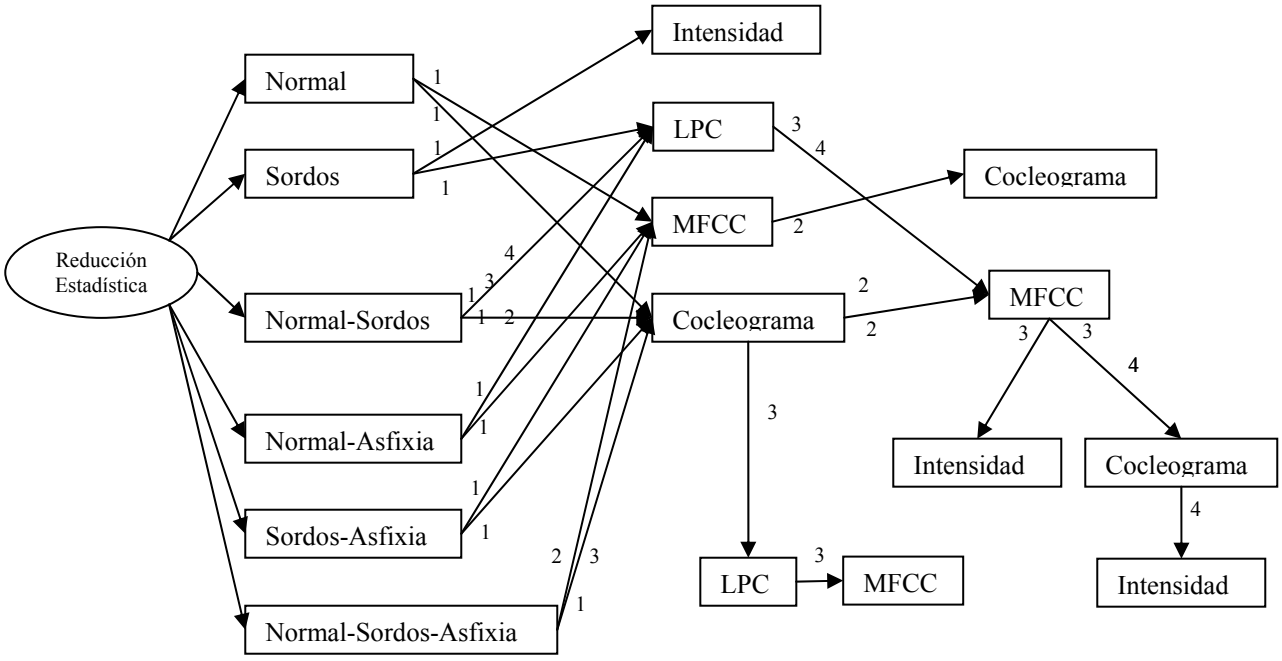


Figura 5.3.1. Diagrama de la base de conocimiento para clasificar llanto de bebé utilizando el método de reducción estadístico.

Los números sobre las transiciones indican el número de características a utilizar y la combinación generada para cada clasificación, dependiendo de la elección y la precisión que se desee, recordando que, entre más características se utilicen para representar al vector característico mayor información tendrá el clasificador para diferenciar entre cada tipo de llanto y mayor será la precisión del clasificador.

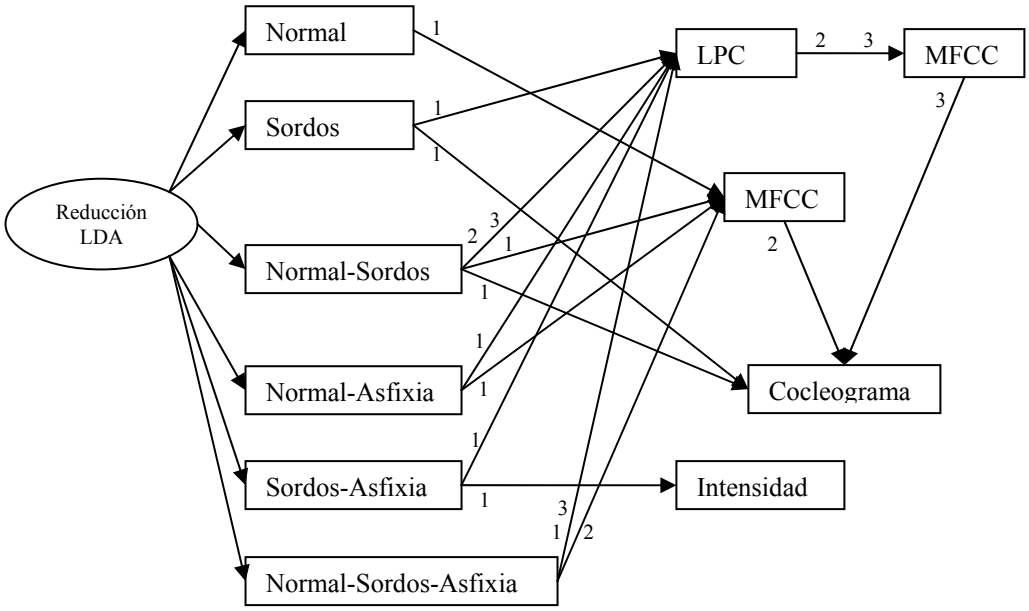


Figura 5.3.2. Diagrama de la base de conocimiento para clasificar llanto de bebé utilizando el método de reducción LDA.

En el caso de las clases “normal-sordos-asfixa”, la combinación de 4 características no mostró un aumento, por lo que se omite dicha combinación, pues el objetivo es mejorar la precisión.

5.4 Comparación con otros trabajos

En el trabajo de Orozco [12] se utilizaron cuatro técnicas de extracción de características: LPC, MFCC, Intensidad y Frecuencia Fundamental. Cada una de las cuales, fue probada de manera independiente con un clasificador basado en redes neuronales. La red neuronal fue entrenada para clasificar llanto de bebés hipoacúsicos y normo-oyentes. Aplicando reducción por Análisis de Componentes Principales (PCA) para un conjunto de datos **A** con 157 muestras de llanto normal (total de muestras de llanto normal) y 157 de llanto hipoacúsico (tomadas aleatoriamente de un conjunto de 879 muestras), aplicando validación cruzada obtuvo los resultados que se muestran en la Tabla 5.4.1.

Tabla 5.4.1. Resultados de Orozco para tres tipos de características y reducción a 50 CP.

50 Componentes Principales	A
LPC (ventana de 50ms)	90.12%
MFCC (ventana de 50ms)	94.58%
Intensidad (ventana de 10ms)	55.01%

Para estas mismas características, aplicando validación cruzada y dos métodos de reducción sobre un conjunto de muestras de 3340 de llanto normal y 2809 de llanto hipoacúsico. Tomando 2809 muestras de cada clase para generar el conjunto de entrenamiento, en este trabajo se obtuvieron los resultados de la Tabla 5.4.2.

Tabla 5.4.2. Resultados obtenidos es este trabajo para tres tipos de características con 5 y 9 atributos.

	Estadística: 5 atributos (ventana de 50ms)	LDA: 9 atributos (ventana de 50ms)
LPC	94.37%	94.15%
MFCC	93.12%	94.71%
Intensidad	92.93%	84.46%

Comparando los resultados obtenidos para LPC, en este trabajo se obtuvo un incremento con ambos métodos de reducción, de aproximadamente 4 puntos porcentuales. Utilizando MFCC, la reducción por LDA obtuvo un valor ligeramente superior, de tan sólo 0.13 décimas. La mayor diferencia se obtuvo para la intensidad

con 29.45 y 37.92 puntos porcentuales de diferencia con ambos métodos de reducción. Para la intensidad, en el trabajo de Orozco se definió una ventana de análisis de 10ms, lo cual pudo haber sido muy pequeña pues para procesamiento de voz, comúnmente se utilizan ventanas de 20-30ms, y en llanto, dada su naturaleza, se toma una ventana aun más grande de 50-100ms.

Orozco llevó a cabo pruebas con un conjunto de entrenamiento de 157 muestras de llanto normal y 879 de llanto hipoacúsico, cada muestra con 304 características LPC, aplicando validación cruzada obtuvo mejores resultados en la clasificación que llegó a un 98.55% de precisión .

En el presente trabajo, utilizando la combinación de cuatro tipos de características y reducción por operaciones estadísticas (5 características de cada tipo), se obtuvo un 98.66% de precisión en las pruebas globales utilizando validación cruzada, y en las pruebas por individuo se alcanzó el 98.94% con reducción estadística y 99.47% con LDA.

A diferencia del trabajo de Orozco, en este trabajo se incrementó el número de muestras de ambas clases, lo cual representa mayores posibilidades de clasificar correctamente muestras totalmente nuevas utilizando un modelo de clasificación ya entrenado. No necesariamente un mayor número de muestras mejorará la precisión del modelo, ya que además se tienen que considerar otros factores como el ruido, la calidad de la grabación, la duración, etc.

Cabe mencionar que el método PCA utilizado en trabajos anteriores [12] y [16] fue probado en este trabajo. Sin embargo, su tiempo de ejecución para procesar las matrices de datos saturó la memoria disponible en diferentes computadoras en las que se realizaron las pruebas. Funcionando perfectamente para conjuntos de datos pequeños, pero no con matrices de mayor tamaño, por lo cual se descartó su uso para ser aplicado sobre el conjunto de datos utilizado en esta tesis.

Capítulo 6

CONCLUSIONES Y TRABAJO FUTURO

6.1 Revisión de objetivos

Haciendo una comparación de los objetivos planteados al inicio de este trabajo de tesis con los resultados obtenidos, tenemos que, para el objetivo general se analizaron cuatro técnicas de extracción de características acústicas aplicadas al llanto de bebés hipoacúsicos y normo-oyentes, incluyendo un método nuevo para extraer cocleogramas. Se generaron diversas caracterizaciones, utilizando los cuatro tipos de características de manera individual y sus combinaciones, las cuales, fueron probadas aplicando un modelo de clasificación basado en redes neuronales, cuyos resultados se analizaron posteriormente, cumpliendo así el objetivo general de este trabajo de investigación.

Dentro de los objetivos particulares, se planteó seleccionar y analizar técnicas de reducción de datos y explorar el uso de operaciones estadísticas para este fin. Dicho objetivo fue alcanzado analizando tres técnicas de reducción, entre ellas las operaciones estadísticas, PCA y LDA. Seleccionando finalmente LDA y operaciones estadísticas, con las cuales se realizaron las pruebas de este trabajo.

Se planteó proponer una mejora a la representación del tipo de llanto mediante la combinación de características en un solo vector, dicho objetivo se cumplió satisfactoriamente, pues se demostró que la combinación de características mejora la representación del llanto de bebé, reduciendo el error y aumentando la tasa de reconocimiento final.

Otro de los objetivos fue seleccionar un modelo de clasificación, para ello, se realizaron diversos experimentos sobre un conjunto de prueba con varios

clasificadores, seleccionando finalmente el que obtuvo los mejores resultados, en este caso, la red neuronal *feed forward*.

Como objetivo particular se planteó definir una base de conocimiento basada en los resultados obtenidos con el fin de servir como referencia futura en la clasificación de diferentes tipos de llanto: normal, sordos y asfixia. Se definieron dos bases de conocimiento, una por cada método de reducción utilizado. Para cada base de conocimiento se consideraron los resultados de las pruebas con cada una de las técnicas de extracción de características, identificando cual de ellas caracterizó mejor cada tipo de llanto. A cada base de conocimiento también se agregaron los resultados de las combinaciones de características, las cuales, se realizaron de manera exhaustiva para las clases normal y sordos, incluyendo además algunas pruebas para la clase asfixia.

Finalmente, uno de los objetivos fue comparar los resultados del modelo de reconocimiento propuesto con los resultados en trabajos anteriores. En este caso, el único trabajo contra el que se realizó una comparación fue el trabajo de Orozco [12], ya que los experimentos en dicho trabajo se realizaron con parte de la base de llantos utilizada en esta tesis, además de que también se utilizaron tres de las técnicas de extracción de características que se prueban en este trabajo de investigación.

De manera general y particular, los objetivos planteados al inicio de este trabajo de investigación fueron cubiertos en su totalidad, en algunos casos, superando las expectativas iniciales, como fue el caso del uso de cocleogramas para caracterizar el llanto y también el uso de operaciones estadísticas como método de reducción.

6.2 Conclusiones

Uno de los objetivos de este trabajo fue explorar el uso de cocleogramas en el proceso de reconocimiento automático del llanto de bebés, en este sentido, de acuerdo a los experimentos realizados se observó que, los cocleogramas igualaron y en algunos casos mejoraron los resultados obtenidos por técnicas como LPC o MFCC de manera individual y en la combinación con otra (s) características.

Además de analizar cada tipo de características de manera individual, se propuso una mejora a la caracterización del llanto basada en la combinación de parámetros. La efectividad de las características acústicas y sus combinaciones se midió de acuerdo al rendimiento de la clasificación, en este sentido, la combinación de características en un solo vector incrementó las tasas de reconocimiento con respecto a los experimentos base con un solo tipo de características, lo cual cumple con los objetivos planteados al inicio de la investigación.

Con respecto a los métodos de reducción, una de las ventajas que presenta el método estadístico es que no requiere información de las clases, ya que cada operación es aplicada de manera individual sobre cada vector característico y el resultado no es variable, es decir, que si se agregan más datos a la matriz, sólo se procesan los datos nuevos y todo el procesamiento anterior se mantiene igual, contrario a lo que sucede con LDA, el cual, al agregar nuevos datos, requiere información de las clases y por tanto llevar a cabo un calculó nuevo de todos los valores de los vectores. Considerando que LDA es una técnica muy utilizada por sus buenos resultados, la reducción por operaciones estadísticas resultó ser un método a considerar en trabajos posteriores, pues en las pruebas globales siempre obtuvo un resultado más alto, superando en este caso a LDA.

En las pruebas por individuo, la combinación LPC-MFCC-Cocleograma con LDA redujo el error a sólo 1 individuo mal clasificado de la clase normal, contra 2 que

obtuvo la reducción estadística, ambos métodos redujeron a 0 el número de errores de la clase sordos.

Se llevaron a cabo diversos experimentos, que en lo posible, abarcaron los diversos escenarios de validación del sistema propuesto, concluyendo en la definición de una base de conocimiento para ayudar a la clasificación de futuras pruebas.

Actualmente en México se están llevando a cabo cambios legislativos que tienen por objeto establecer la realización de pruebas de audición en todos los recién nacidos, con el fin de identificar los posibles casos de hipoacusia y de realizar sobre éstos un seguimiento continuo del desarrollo de la capacidad auditiva. Sin embargo, la aplicación de estas medidas se ve muy limitada por las restricciones presupuestarias, por lo que las pruebas se limitan, en muchos centros hospitalarios, a los recién nacidos que pertenecen a grupos de riesgo y en la actualidad no se realizan de forma generalizada en nuestro país. Por lo cual, sería de gran utilidad contar con una herramienta auxiliar, no invasiva, que ayude al médico a un diagnóstico precoz en este tipo de patologías, sobre todo, para ser aplicada en aquellas zonas donde no se cuente con médicos especialistas.

6.3 Trabajo futuro

Las líneas de investigación futuras estarán enfocadas a la detección de nuevas patologías como la hiperbilirrubinemia, hidrocefalia, hipoglucemia, etc.

El diseño y desarrollo de un sistema que implemente la base de conocimiento definida en este trabajo se plantea como un trabajo futuro. Dicho sistema deberá considerar la evolución y futuras adaptaciones a las necesidades que surjan de la investigación, como la inclusión de nuevas clases.

Se pretende aplicar el modelo propuesto para identificar subtipos de llanto normal, como hambre y dolor, y definir las características o combinaciones que mejor representen cada tipo de llanto para ser agregadas a la base de conocimiento.

En el caso de la clase asfixia es necesario colectar suficientes muestras para balancear las clases y reducir el error en la clasificación de este tipo de llanto.

Aplicar nuevos métodos de análisis a la onda de llanto, pues se considera que existe mayor información en la onda de llanto que puede ser útil para saber más sobre el estado físico y de salud del bebé.

Los cocleogramas mostraron buenos resultados en la clasificación. Sin embargo, creemos que aun contiene información útil que puede ser aprovechada en la clasificación del llanto.

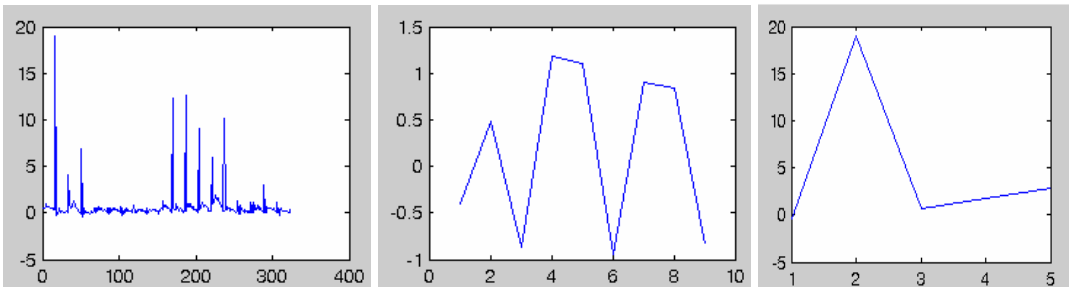
La detección de grados de sordera es importante ya que en el diagnóstico es necesario disponer de un sistema de referencia que permita generar diversas alternativas de tratamiento y excluir otras. En este sentido, la colección de muestras bien etiquetadas permitirá llevar a cabo este tipo de clasificación, pues al igual que en el caso del llanto de asfixia se cuenta actualmente con pocas muestras con la información sobre la medición auditiva de cada oído incluida. Por lo que se requieren más muestras con este tipo de información.

Seguir creciendo la base de conocimiento para identificar más tipos de llanto. En el caso de la identificación de patologías, complementar la información con el desarrollo de un sistema experto difuso para auxiliar al médico en el diagnostico y tratamiento recomendados.

APÉNDICE A

En este apéndice se muestran las representaciones gráficas de los vectores característicos y su transformación al ser reducidos por los métodos LDA y operaciones estadísticas.

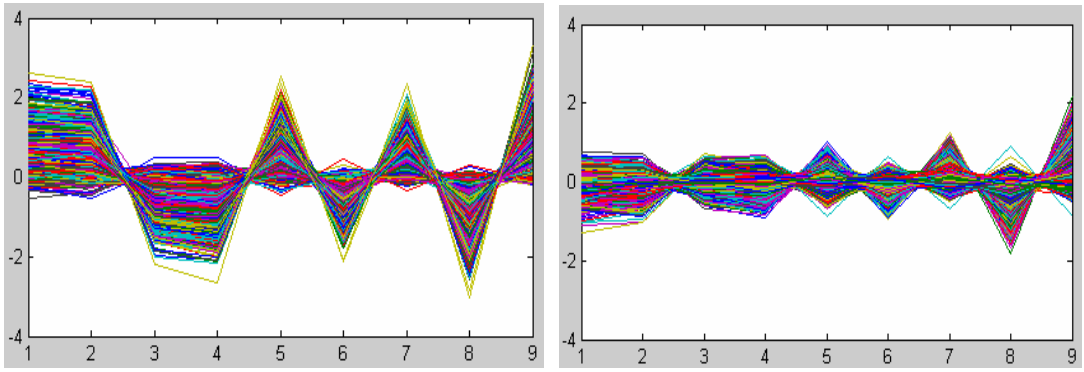
LPC



(a) (b) (c)

Figura A1. (a) Vector de características LPC; (b) Reducción por 9 LDA's; (c) Reducción por operaciones estadísticas.

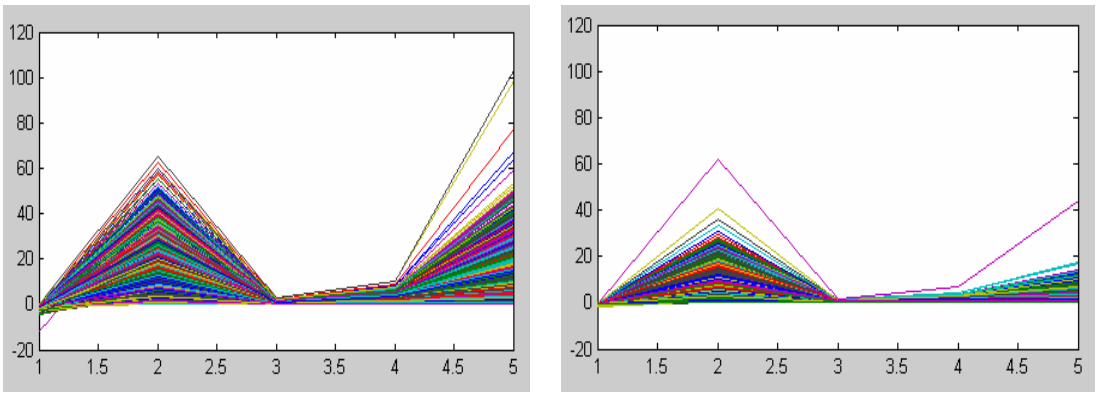
Reducción de todo el espacio por LDA de vectores característicos LPC:



(a) Normal (b) Sordos

Figura A2. Reducción de todo el espacio por LDA de vectores característicos LPC. (a) Clase Normal, (b) Clase Sordos

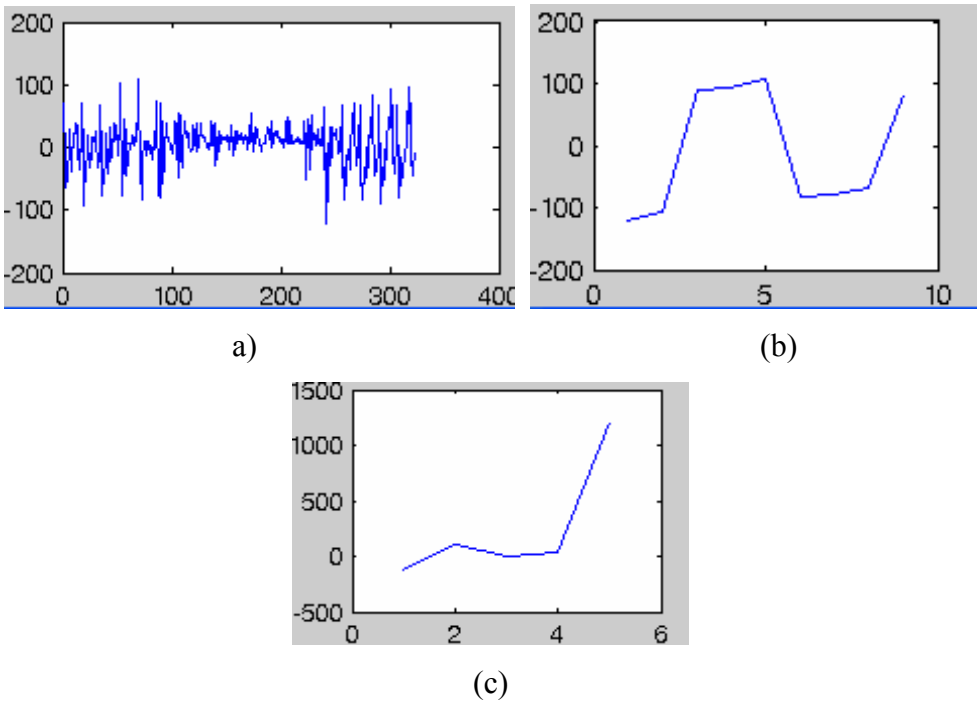
Reducción de todo el espacio por operaciones estadísticas de vectores característicos LPC:



(a) Normal (b) Sordos

Figura A3. Reducción de todo el espacio por operaciones estadísticas de vectores característicos LPC. (a) Clase Normal, (b) Clase Sordos

MFCC

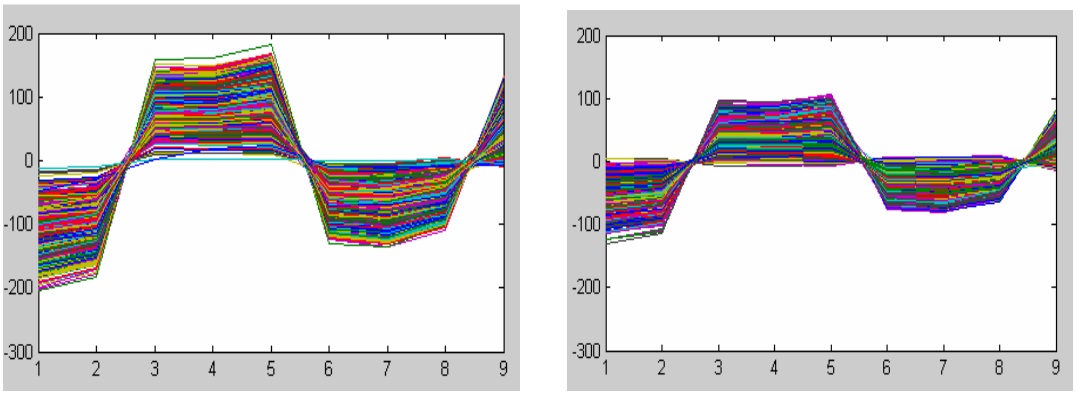


a) (b)

(c)

Figura A4. (a) Vector de características MFCC; (b) Reducción por LDA; (c) Reducción por operaciones estadísticas

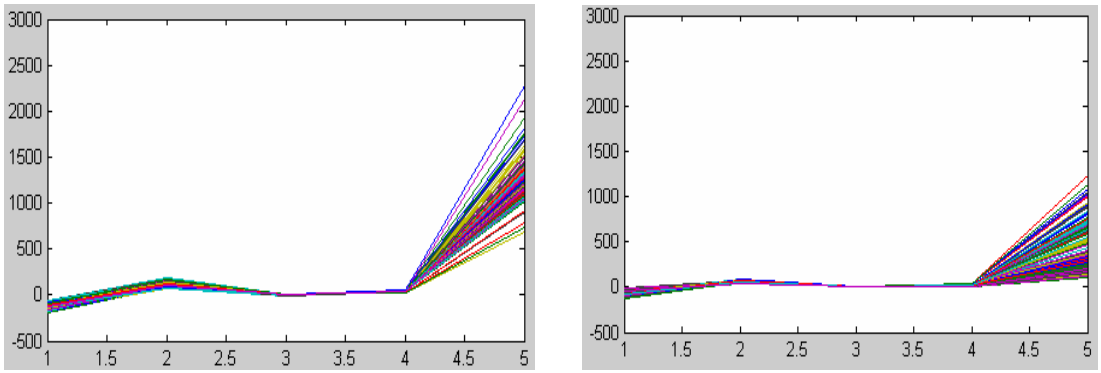
Reducción de todo el espacio por 9 LDA de vectores característicos MFCC:



(a) Normal (b) Sordos

Figura A5. Reducción de todo el espacio por LDA de vectores característicos MFCC. (a) Clase Normal, (b) Clase Sordos

Reducción del espacio por operaciones estadísticas de vectores característicos MFCC:



(a) Normal (b) Sordos

Figura A6. Reducción de todo el espacio por operaciones estadísticas de vectores característicos MFCC. (a) Clase Normal, (b) Clase Sordos

COCLEOGRAMA

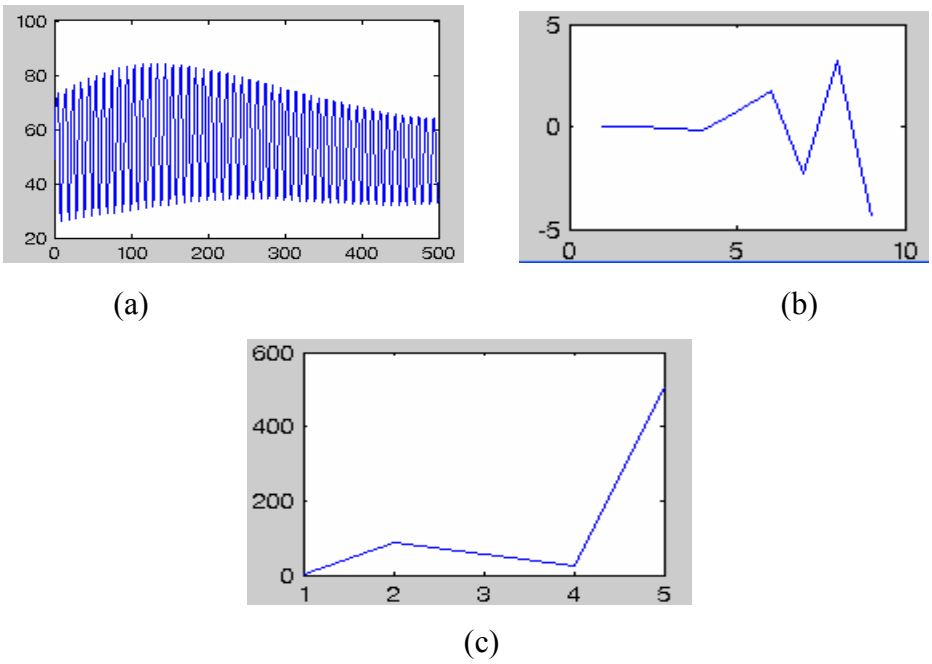


Figura A7. (a) Vector de características del cocleograma; (b) Reducción por LDA; (c) Reducción por operaciones estadísticas

Reducción de todo el espacio por 9 LDA's de vectores característicos del cocleograma:

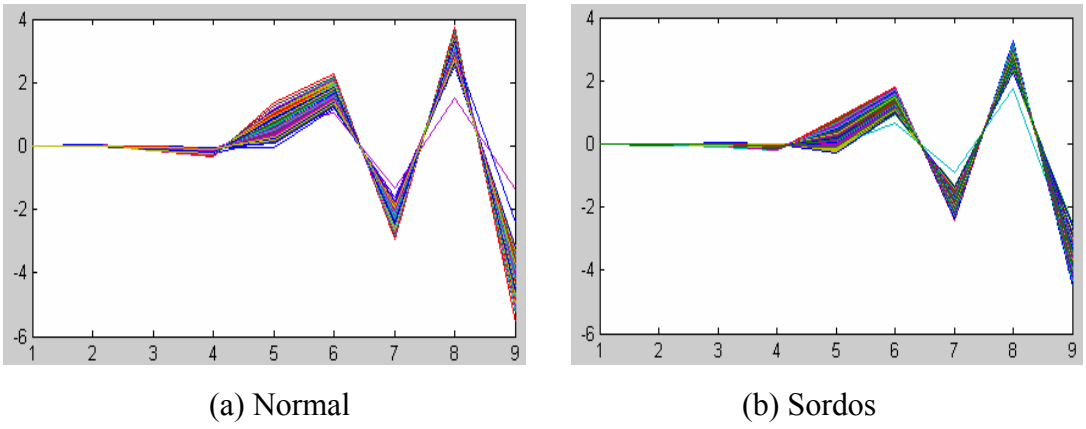
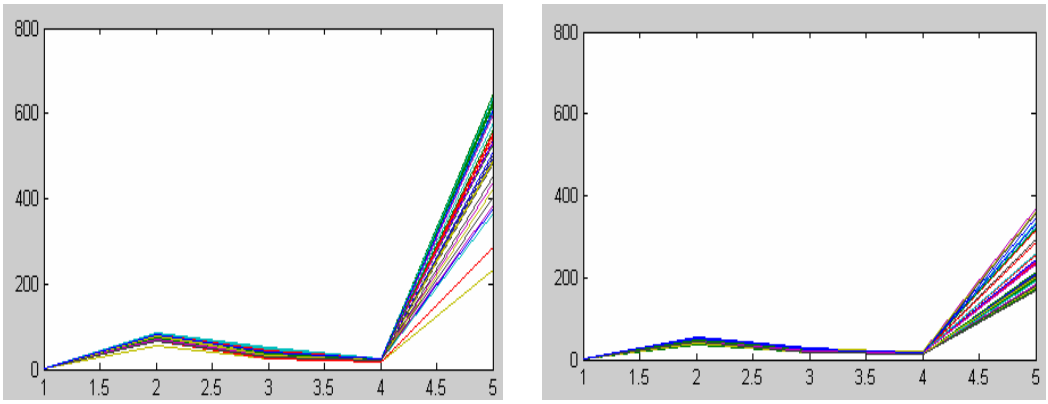


Figura A8. Reducción de todo el espacio por LDA de vectores característicos del cocleograma. (a) Clase Normal, (b) Clase Sordos

Reducción del espacio por operaciones estadísticas de vectores característicos del cocleograma:



(a) Normal (cocleograma) (b) Sordos (cocleograma)

Figura A9. Reducción del espacio por operaciones estadísticas de vectores característicos del cocleograma. (a) Clase Normal, (b) Clase Sordos

APÉNDICE B

La Tabla B1, muestra el resumen de los mejores resultados obtenidos en las pruebas globales (validación cruzada) utilizando vectores con una sola característica.

Tabla B1. Resumen de los mejores resultados obtenidos en las pruebas globales considerando cuatro tipos de características: LPC, MFCC, Intensidad y Clocleograma.

Tipo de llanto a clasificar	Característica Sugerida	% Pruebas Globales
Normal-sordos (R. Estadística)	LPC	94.37%
	Cocleograma	94.02%
Normal-sordos (R. LDA)	LPC	94.15%
	MFCC	94.71%
	Cocleograma	95.69%
Normal-asfixia (R. Estadística)	MFCC	91.5%
	LPC	91.87%
Normal-asfixia (R. LDA)	MFCC	89.5%
	LPC	91.87%
Sordos-asfixia (R. Estadística)	Cocleograma	88.75%
	MFCC	87.5%
Sordos-Asfixia (R. LDA)	LPC	68.75%
	Intensidad	70%
Normal-sordos-asfixia (R. Estadística)	Cocleograma	90%
Normal-sordos-asfixia (R. LDA)	LPC	75.83%

Un resumen de los mejores resultados en las pruebas globales con las combinaciones de dos, tres y cuatro tipos de características se muestra en la Tabla B2.

Tabla B2. Resumen de los mejores resultados obtenidos por la combinación de dos, tres y cuatro tipos de características en las pruebas globales, con validación cruzada.

Tipo de llanto a clasificar	Combinación dos características	% Pruebas Globales
Normal-sordos (R. Estadística)	MFCC-Cocleograma	97.58%
Normal-sordos (R. LDA)	LPC-MFCC	96.61%
Normal-Sordos-Asfixia (R. Estadística)	MFCC-Cocleograma	92%
Normal-Sordos-Asfixia (R. LDA)	MFCC-Cocleograma	84.69%
	Combinación tres características	
Normal-sordos (R. Estadística)	LPC-MFCC-Intensidad	98.30%
Normal-Sordos (R. LDA)	LPC-MFCC-Cocleograma	96.69
Normal-Sordos-Asfixia (R. Estadística)	LPC-MFCC-Cocleograma	92.8%
Normal-Sordos-Asfixia (R. LDA)	LPC-MFCC-Cocleograma	84.83%
	Combinación cuatro características	
Normal-sordos (R. Estadística)	LPC-MFCC-Cocleograma-Intensidad	98.66%,

La Tabla B3, muestra el resumen de las características que mejor clasifican al tipo de llanto normal y sordo de acuerdo a los resultados de las pruebas por individuo.

Tabla B3. Resumen de los mejores resultados para clasificar llanto de tipo normal y sordo, utilizando un tipo de características en las pruebas por individuo.

Tipo de llanto a clasificar	Características con mejor resultado	% de precisión por clase
Normal (R. Estadística)	Cocleograma	95.93%
	MFCC	95.12%
Normal (R. LDA)	MFCC	95.93%
	LPC	91.87%
Sordos (R. Estadística)	LPC	100% (ambas)
	Intensidad	
Sordos (R. LDA)	LPC	100% (ambas)
	Cocleograma	

En las pruebas por individuo las combinaciones que arrojaron los mejores resultados se presentan en la Tabla B4.

Tabla B4. Resumen de los mejores resultados obtenidos por la combinación de características en las pruebas por individuo para clasificar las clases "normal-sordos".

Tipo de llanto a clasificar	Combinación dos tipos de características	% precisión por clase
Normal-sordos (R. Estadística)	MFCC-Cocleograma	Normal: 95.93% Sordos: 100%
	LPC-Cocleograma	Normal: 96.75% Sordos: 100%
Normal-sordos (R. LDA)	LPC-MFCC	Normal: 98.37% Sordos: 100%
	MFCC-Cocleograma	Normal: 95.12% Sordos: 100%
	Combinación tres tipos de características	
Normal-sordos (R. Estadística)	LPC-MFCC-Intensidad	Normal: 97.56% Sordos: 100%
	LPC-MFCC-Cocleograma	Normal: 97.56% Sordos: 100%
Normal-sordos (R. LDA)	LPC-MFCC-Cocleograma	Normal: 99.19% Sordos: 100%
	Combinación cuatro tipos de características	
Normal-sordos (R. Estadística)	LPC-MFCC-Cocleograma-Intensidad	Normal: 98.37% Sordos: 100%

FIGURAS

TABLAS

BIBLIOGRAFÍA

[1] Wasz-Hockert O., Partanen T., Vuorenkoski V., Valanne E. and Michelsson K.; "The identification of some specific meanings in infant vocalization"; Experientia, Vol. 20, pp. 154-156, 1964.

[2] Wasz-Hockert O., Vuorenkoski V., Lind J., Partanen T. and Valanne E.; *The infant cry: A spectrographic and auditory analysis;* Spastics International Medical Publication, Lavenham, U.K. 1968.

[3] Sokol-Hyde, "Evaluación auditiva", Pediatrics in review en español; Editorial Médica AWWE, American Academy of Pediatrics, Volumen 23 No.8 pp. 283-289, 2002.

[4] Rodríguez J.A, Chavira C. L., Montes de Oca E.; "Frecuencia de defectos auditivos en 16 estados de México / Frecuency of hearing impairment in 16 states of Mexico"; Anales de otorrinolaringología mexicana / Sociedad Mexicana de Otorrinolaringología.; Vol. 46(3): pp.115-117, 1991.

[5] Hospital General de México, "Informe Enero-Junio de 2003"; Servicio de Audiología y Foniatría Sección 104, Clínica para la detección de problemas auditivos, 2003. disponible en web:
http://hgm.salud.gob.mx/pdf/servicios/cliproau.pdf

[6] Martínez C.F., Jara N. "Análisis espectrofonográfico del llanto en recién nacidos de término con riesgo neurológico"; Instituto Nacional de Perinatología; Publicación del Hospital Infantil de México Federico Gómez, Instituto Nacional de Salud, Publicación Bimestral, Vol. 63, Sup 1, No. 1, México D.F. Enero–Febrero 2006. disponible en web:
http://www.medigraphic.com/pdfs/bmhim/hi-2006/his061.pdf

[7] Arch-Tirado E., Mandujano M, García-Torices L, Martínez-Cruz et al, "Análisis del llanto del niño hipoacúsico y del niño normo-oyente", Cirugía Cirujano, 72 (4): pp. 271-276, 2004.

[8] Cohen A. and Zmora E.; "Automatic classification of infants' hunger and pain cry"; In Proc. Int. Conf. Digital Signal Processing, Cappellini, V. and Constantinides, A.G., Eds., Elsevier, Amsterdam, pp. 667-672, 1984.

[9] Petroni M., Malowany A., Johnston C., Stevens B.; "Identification of Pain from Infant Cry Vocalizations Using Artificial Neural Networks (ANNs)"; The International Infant Cry Research Group. Applications and Science of Artificial Neural Networks, The International Society for Optical Engineering. Volumen 2492. pp.729-738, 1995.

[10] Ekkel T.; "Neural Network-Based Classification of Cries from Infants Suffering from Hypoxia-Related CNS Damage"; Tesis de Maestría, University of Twente, The Netherlands, 2002.

[11] Lederman D.; "Automatic Classification Of Infant's Cry"; Tesis de Maestría, Ben-Gurion University of the Negev, Israel, 2002.

[12] Orozco G. J., Reyes C.A.; "Extracción de Análisis de Características Acústicas del Llanto de Bebés para su Reconocimiento Automático Basado en Redes Neuronales"; Tesis de Maestría, Instituto Nacional de Astrofísica Óptica y Electrónica, Puebla, México, 2003.

[13] Cano S. D., Escobedo D. I.; "Clasificación de Unidades de Llanto Infantil Mediante el Mapa Auto-Organizado de Kohonen"; I Taller AIRENE sobre Reconocimiento de Patrones con Redes Neuronales, Universidad Católica del Norte, pp 24-29; Chile, 1999.

[14] Cano S. D., Escobedo D. I., Regueiferos L., Capdevila L.; "15 Años del Cry Analysis en Cuba: Resultados y Perspectivas"; VI Congreso Internacional de Informática en Salud, Santiago de Cuba, 2007.

[15] Shamma S., Byrne W., Robinson J.; "The Auditory Processing and Recognition of Speech"; Association for Computational Linguistics, 1989.

[16] Barajas S. E., Reyes C.A; "Clasificación de Llanto Infantil"; Tesis de Maestría, Instituto Nacional de Astrofísica Óptica y Electrónica, Puebla, México, 2006.

[17] Sibbald A.; "Diagnostico de hipoacusia"; Sociedad Argentina de Pediatría, PRONAP, pp. 37-50, Argentina, 1996.

[18] Proakis J.G.,Manolakis D.G.; *Tratamiento Digital de Señales, Principios, algoritmos y aplicaciones*; 3ra Edición, Prentice Hall, Cap. 1. pp. 1-38, 1998

[19] Jayant N.S., Noll P.; *Digital Coding of Waveforms, principles and applications to speech and video*; Prentice-Hall, 1984.

[20] Castañeda P. F.; *El lenguaje verbal del niño*; Universidad Nacional Mayor de San Marcos (NMSM), ISBN: 9972-46-073-8, Lima Perú, pp.125-140, 1999.

[21] Luria A.R.; *El cerebro en acción*; Ed. Fontanella, Barcelona 1974.

[22] Britos P., Hossian A., García-Martínez R., Sierra E.; "Minería de datos basada en sistemas inteligentes: Redes Neuronales Artificiales"; Cap 7 Nueva Librería, Buenos Aires, Argentina., 2005.

[23] Gerstner W.; *Supervised learning for neural networks: A tutorial with Java Exercises.*, 2002., disponible en web: http://lcn.epfl.ch/tutorial/docs/supervised.pdf

[24] Deller J. R., Proakis J. G. and Hansen J. H. L.; "Discrete-Time Processing of Speech Signals"; Mac Millan, N. Y., 1993.

[25] Boersma P. & Weenink D; PRAAT: Doing phonetics by computer (Versión 4.4.3.1) [programa computacional]. Obtenido en Agosto 2006, del sitio web: www.praat.org.

[26] Platt J.C.; "Fast Training of Support Vector Machines using Sequential Minimal Optimization"; Microsoft Research, EUA ,2000.

[27] Quinlan R.; "C4.5: Programs for Machine Learning"; Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[28] Breiman L.; "Random Forests"; *Machine Learning*; Cap 45 (1):pp.5-32, 2001.

[29] Mitchell T; "Bayesian Learning"; *Machine Learning*, McGraw Hill, Cap 6. 1997.

[30] Bauer E., Kohavi R.; "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants"; *Machine Learning*, Kluwer Academic Publishers, Boston, Cap 36: pp.105-142, 1999.

[31] Wolpert H. D.; "Stacked generalization"; *Neural Networks*, Pergamon Press Cap 5: pp.241-259, 1992.

[32] Reyes-Galaviz O.F., Arch-Tirado E., Reyes-García C.A.; "Classification of Infant Crying to Identify Pathologies in Recently Born Babies with ANFIS"; Computers Helping People with Special Needs 9th International Conference, ICCHP 2004 Paris, France, pp 408-415, 2004.

[33] Reyes-Galaviz O. F., Reyes-García C.A.; "Infant Cry Classification to Identify Hypoacoustics and Asphyxia with Neural Networks"; Lecture Notes in Computer Science Vol. 2972 Springer, Berlin, pp 69-78, ISBN 3-540-21459-3, ISSN 0302-9743. 2004.

[34] Hagan M. T. and Menhaj M.; "Training feedforward networks with the Marquardt algorithm"; IEEE Transactions on Neural Networks, vol. 5, no. 6, pp. 989-993, 1994.

[35] Balakrishnama S., Ganapathiraju A.; "Linear Discriminant Analysis - A brief tutorial"; Institute for Signal and Information Processing Department of Electrical and Computer Engineering Mississippi State University.

[36] Amaro-Camargo E., Reyes-García C.A.; "Applying Statistical Vectors of Characteristics and Ensembles for the Automatic Recognition of Infant Cry"; Lecture Notes in Computer Sciences (LNCS) 4681, edited by De-Shuang Huang, Laurent Heutte and Marco Loog, Springer, Berlin, pp. 1078-105, ISBN: 978-3-540-74170-4, ISSN: 0302-9743., 2007.

[37] Rosete D. M.; "Es curable la sordera?"; artículo disponible en web: http://www.mipediatra.com/infantil/sordera2.htm, actualizado en el 2006.

[38] Haykin S., "Neural Networks: A Comprehensive Approach", IEEE Computer Society Press, Piscataway, USA (1994).

[39] World Health Organization (WHO); “Primary Ear And Hearing Care Training Resource”; Chronic Disease Prevention and Management, WHO Library Cataloguing-in-Publication Data, Switzerland ,2006. http://whqlibdoc.who.int/publications/2006/9241592710_eng.pdf