

**SISTEMA DE CANCELACIÓN DE SEÑALES DE AUDIO EMITIDAS POR UN  
INDIVIDUO EN SUS ÁREAS CIRCUNDANTES**

**ADRIANA ROCÍO GARCÍA FANDIÑO  
JULY PAOLA LICONA BLANCO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMATICA  
BUCARAMANGA  
2008**

**SISTEMA DE CANCELACIÓN DE SEÑALES DE AUDIO EMITIDAS POR UN  
INDIVIDUO EN SUS ÁREAS CIRCUNDANTES**

**ADRIANA ROCÍO GARCÍA FANDIÑO  
JULY PAOLA LICONA BLANCO**

**Trabajo de Investigación de pregrado de Ingeniería de Sistemas e  
Informática**

**Director:  
Mpe. Henry Arguello Fuentes  
Profesor  
Escuela de Ingeniería de Sistemas e Informática**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2008**

*A Dios, Motor principal de mi vida amigo incondicional y padre excepcional.*

*A mis papás Álvaro García y Mariela Fandiño, por el esfuerzo, los consejos, el amor  
y por ser un gran ejemplo de vida; gran cariño y admiración.*

*A mis hermanos, Daniel y Mauricio, personas importantes en mi vida, los quiero mucho.*

*A mis amigos, de toda la vida y los que he ido conociendo en el transcurrir de los años,  
Personas que estuvieron y estarán siempre para dar una voz de aliento, una sonrisa amable,  
una mano de ayuda y un consejo apropiado.*

*A todos los profesores por sus valiosas enseñanzas, siempre hay mucho que aprender y  
agradecer.*

*A mis compañeros de la Universidad por las tardes de estudio, el intercambio de  
experiencias y anécdotas que hicieron la vida universitaria una agradable experiencia de  
vida.*

*A toda la gente que ha pasado por mi vida y con quienes compartí en algún momento.*

*Adri*

*A Dios, por su voluntad y misericordia estoy aquí y ahora. Porque todo lo puedo en  
Cristo que me fortalece.*

*A mis padres, Ada Blanco y Edgar Licona, mi apoyo incondicional en el camino.  
Por todos los sacrificios, la confianza permanente, el amor y las enseñanzas recibidas.*

*A mi mamá, Carlina Castilla de Licona (q.e.p.d). Por todos los años que me regaló.  
Por todo el amor. Por su crianza. Por ser una mujer tan especial y dispuesta a dar todo.  
Porque no hay un solo día en que no la recuerde y extrañe*

*A mis hermanos, que los amo. Por todos los instantes compartidos y por los que deje de  
estar por perseguir esta meta.*

*A mi familia, que ocupan un lugar muy importante en mi corazón, siempre preocupados por  
mi bienestar y porque siempre tuvieron la palabra indicada en el momento preciso.*

*A mis amigos, el apoyo cuando me sentí desvanecer y mis compañeros en cada aventura.  
Por las lágrimas que secaron y por las mil sonrisas que me regaron.*

*A todas las personas que me han regalado su tiempo, sus historias, enseñanzas, su cariño y  
paciencia.*

*July Pao*

## **AGRADECIMIENTOS**

Los autores del libro expresan sus agradecimientos a:

A Dios y nuestra familia que son los fundamentos de nuestra vida, por el apoyo, confianza y amor incondicional.

Al profesor HENRY ARGUELLO FUENTES, Magíster en Potencia Eléctrica. Profesor de la Universidad Industrial de Santander. Por sus aportes como Director de este trabajo de grado, poniendo a nuestra disposición su experiencia y conocimiento, por sus enseñanzas, tiempo y paciencia, entrega profesional y acertada dirección.

A los profesores del programa de Ingeniería de Sistemas e Informática por todos los conocimientos transmitidos y por la vocación de la enseñanza algunas veces desagradada. En especial al profesor Víctor E. Martínez Abaunza, por su colaboración y apoyo en el trabajo en el grupo de investigación GIIB.

Al LABORATORIO DEL MEDIO AMBIENTE de Bucaramanga, en especial al Técnico Ambiental FABIO GONZALEZ, por la buena disposición y paciencia en el desarrollo de las pruebas donde era necesario el uso del sonómetro que nos facilitaron.

Al Director General del CTI, OSCAR FRANCISCO CASTELLANOS PRADA, por su amabilidad y buena voluntad en el préstamo del laboratorio de acústica del CTI.

Al CENTIC, en especial a las personas encargadas de la sala de grabación, por permitirnos utilizar el espacio para las pruebas y equipos especiales para la grabación de señales de voz necesarias en nuestro proyecto.

Al grupo de investigación GIIB y sus integrantes, por el apoyo y aportes en el desarrollo de nuestro trabajo de grado y el crecimiento a nivel personal, destacando la importancia del trabajo colaborativo y los seminarios de formación.

A nuestros amigos “los 2030”, los que siempre serán “los de cuarto” y a los encontrados en el transcurrir de la vida universitaria. Por su agradable compañía,

las tardes de estudio y de ocio, los consejos, las voces de aliento, el consejo apropiado, los chistes, por aguantarnos y por todas las anécdotas que quedaran en la memoria y le contaremos a nuestros futuros hijos orgullosos de sus madres.

Por último, pero sin restar importancia. A la Universidad Industrial de Santander, nuestra alma mater. Por habernos abierto la puerta y habernos dotado de conocimiento y sabiduría de tal forma que han transformado nuestra forma de pensar. Por ser el punto de encuentro de nuevas amistades que no olvidaremos.

## TABLA DE CONTENIDO

<b>INTRODUCCIÓN</b>	<b>19</b>
<b>1. DEFINICIÓN DEL PROBLEMA</b>	<b>4</b>
1.1 ANTECEDENTES	4
1.2 FORMULACIÓN	6
1.3 DESCRIPCIÓN	6
<b>2 JUSTIFICACIÓN</b>	<b>8</b>
<b>3 OBJETIVOS</b>	<b>9</b>
3.1 OBJETIVO GENERAL	9
3.2 OBJETIVOS ESPECÍFICOS	9
<b>4 MARCO TEÓRICO</b>	<b>10</b>
4.1 COEFICIENTES DE PREDICCIÓN LINEAL LPC [6][13]	10
4.1.1 Definición	10
4.1.2 Principios básicos del análisis de predicción	11
4.2 REDES NEURONALES ARTIFICIALES	17
4.2.1 Conceptos [9]	17
4.2.2 Modelo Neuronal [4]	19
4.3 ONDAS SONORAS [2]	20

<b>4.3.1</b>	<b>Interferencia de Ondas Sonoras</b>	<b>21</b>
<b>4.3.1.1</b>	<b>Interferencia Constructiva</b>	<b>21</b>
<b>4.3.1.2</b>	<b>Interferencia Destructiva</b>	<b>22</b>
<b>4.3.2</b>	<b>Frecuencia</b>	<b>23</b>
<b>4.3.3</b>	<b>Amplitud</b>	<b>24</b>
<b>4.3.4</b>	<b>Potencia Sonora</b>	<b>25</b>
<b>4.3.5</b>	<b>Intensidad Sonora</b>	<b>26</b>
<b>4.3.6</b>	<b>Nivel de Presión Acústica</b>	<b>27</b>
<b>4.4</b>	<b>SENSIBILIDAD HUMANA</b>	<b>30</b>
<b>4.4.1</b>	<b>Relación Señal/Ruido</b>	<b>31</b>
<b>5</b>	<b>ESTADO DEL ARTE</b>	<b>32</b>
<b>5.1</b>	<b>PREDICCIÓN DE SEÑALES DE VOZ</b>	<b>32</b>
<b>5.1.1</b>	<b>Coeficientes de Predicción Lineal</b>	<b>32</b>
<b>5.1.2</b>	<b>Predicción con Redes Neuronales</b>	<b>33</b>
<b>5.2</b>	<b>CANCELACIÓN DE SEÑALES</b>	<b>35</b>
<b>5.3</b>	<b>TECNOLOGÍAS QUE ABORDAN LA PROBLEMÁTICA</b>	<b>36</b>
<b>6</b>	<b>DESARROLLO DEL PROYECTO</b>	<b>37</b>
<b>6.1</b>	<b>CAPTURA DE LA SEÑAL</b>	<b>37</b>
<b>6.1.1</b>	<b>Recursos empleados en la captura</b>	<b>38</b>

<b>6.1.2</b>	<b>Descripción del ambiente</b>	<b>39</b>
<b>6.2</b>	<b>PREDICCIÓN DE LA SEÑAL</b>	<b>40</b>
<b>6.2.1</b>	<b>Predicción de voz mediante Coeficientes de Predicción Lineal</b>	<b>41</b>
<b>6.2.1.1</b>	<b>Procedimiento para la predicción</b>	<b>41</b>
<b>6.2.1.2</b>	<b>Resultado del Método de Predicción Lineal</b>	<b>44</b>
<b>6.2.2</b>	<b>Predicción de la señal de voz mediante Redes Neuronales Artificiales</b>	<b>48</b>
<b>6.2.2.1</b>	<b>Elección de la arquitectura adecuada</b>	<b>50</b>
<b>6.2.2.2</b>	<b>Optimización de la Red</b>	<b>56</b>
<b>6.3</b>	<b>RESULTADOS COMPARATIVOS DE LAS TÉCNICAS DE PREDICCIÓN</b>	<b>59</b>
<b>6.3.1</b>	<b>Análisis de Frecuencia</b>	<b>62</b>
<b>6.3.2</b>	<b>Pruebas Humanas</b>	<b>64</b>
<b>6.4</b>	<b>CANCELACIÓN DE LA SEÑAL</b>	<b>65</b>
<b>6.4.1</b>	<b>Pruebas numéricas de cancelación mediante un Sonómetro</b>	<b>66</b>
<b>6.4.1.1</b>	<b>Escoger la mejor posición del parlante</b>	<b>68</b>
<b>6.4.1.2</b>	<b>Posición del Oyente Externo</b>	<b>71</b>
<b>6.4.1.3</b>	<b>Ambiente para la toma de muestras</b>	<b>74</b>
<b>6.4.1.4</b>	<b>Medidas a tomar para el experimento</b>	<b>75</b>
<b>6.4.2</b>	<b>Resultados del experimento</b>	<b>75</b>

<b>6.4.2.1</b>	<b>Escenario 1: Ruido ambiental a 46 db</b>	<b>75</b>
<b>6.4.2.2</b>	<b>Escenario 2: Ruido ambiental a 33 dB.</b>	<b>80</b>
<b>6.4.2.3</b>	<b>Resultados finales de las pruebas numéricas</b>	<b>84</b>
<b>6.4.3</b>	<b>Pruebas cualitativas de cancelación con usuario final</b>	<b>86</b>
<b>6.5</b>	<b>PREDICCIÓN Y CANCELACIÓN DE LA SEÑAL</b>	<b>88</b>
<b>6.6</b>	<b>INTERFAZ GRÁFICA</b>	<b>96</b>
<b>7</b>	<b>CONCLUSIONES Y TRABAJOS FUTUROS</b>	<b>107</b>
<b>7.1</b>	<b>CONCLUSIONES</b>	<b>107</b>
<b>7.2</b>	<b>TRABAJOS FUTUROS</b>	<b>109</b>
	<b>BIBLIOGRAFÍA</b>	<b>111</b>
	<b>ANEXO A Pruebas de algoritmos de entrenamiento</b>	<b>114</b>
	<b>ANEXO B Pruebas de cancelación escenarios 1 y 2</b>	<b>116</b>
	<b>ANEXO C Potencias de cancelación cti sin ruido ambiental</b>	<b>118</b>
	<b>ANEXO D Pruebas cualitativas con usuario final</b>	<b>119</b>
	<b>ANEXO E Cancelación de señales con técnicas de predicción</b>	<b>123</b>
	<b>ANEXO F Potencia de cancelación de señales con tecnicas sin ruido ambiental</b>	<b>126</b>
	<b>ANEXO G Trabajo futuro: DPS</b>	<b>126</b>
	<b>ANEXO H Artículo predicción</b>	<b>126</b>

## LISTA DE TABLAS

Tabla 1. Niveles sonoros y respuestas humanas.....	29
Tabla 2. Estándares máximos permisibles de niveles de emisión de ruido expresados en decibeles DB(A).....	30
Tabla 3. Especificaciones técnicas del computador.....	38
Tabla 4. Descripción de la tarjeta de audio.....	38
Tabla 5. Micrófono de escritorio omnidireccional Genius MIC-01 <sup>a</sup> .....	39
Tabla 6. Micrófono Cobalt CO 11.....	39
Tabla 7. Características de la sala de grabación del CENTIC.....	40
Tabla 8. Pruebas de variación de parámetros con LPC.....	44
Tabla 9. Elección de la mejor arquitectura.....	50
Tabla 10. Elección del mejor algoritmo de entrenamiento.....	57
Tabla 11. Resultados de predicción por LPC y RNA al variar los individuos y las frases.....	61
Tabla 12. Pruebas Humanas de Predicción con frases repetitivas.....	64
Tabla 13. Especificaciones técnicas del sonómetro.....	67
Tabla 14. Potencia según la disposición de los parlantes.....	70
Tabla 15. Mediciones de potencia para el plano de 0 a 200 cm modificando la posición cada 50 cm, ruido ambiental promedio de 46 dB.....	76
Tabla 16. Mediciones de potencia para el plano de 0 a 132 cm con variación de 33 cm en cada punto, promedio de ruido ambiental 33 dB.....	80
Tabla 17. Mediciones de potencia para el plano de 0 restando el ruido ambiental.....	85
Tabla 18. Mediciones de potencia de cancelación con técnicas.....	89
Tabla 19. Potencia atenuada en dB de las tres formas de cancelación.....	91
Tabla 20. Potencias sin ruido ambiental.....	94

## LISTA DE FIGURAS

Figura 1.	Modelo de producción de síntesis de voz	11
Figura 2.	Interferencia constructiva de dos pulsos	22
Figura 3.	Interferencia destructiva de dos pulsos	22
Figura 4.	Sala de Grabación CENTIC	40
Figura 5.	Diagrama de flujo del algoritmo de predicción por LPC	43
Figura 6.	Gráfica segmento de entrada y muestras predichas por LPC	43
Figura 7.	Gráfica del error variando p y M para elegir la mejor configuración	46
Figura 8.	Señal original correspondiente a la frase "Aló, Buenos días"	47
Figura 9.	Señal predicha con LPC correspondiente a la frase "Aló, Buenos días"	47
Figura 10.	Diagrama de flujo del algoritmo de predicción con RNA	49
Figura 11.	Gráfica segmento de entrada y muestras predichas por RNA	49
Figura 12.	Resultado de la función de rendimiento para las diferentes arquitecturas	52
Figura 13.	Resultado de los aciertos para las diferentes arquitecturas	53
Figura 14.	Resultado de la generalización para las diferentes arquitecturas	54
Figura 15.	Resultado de los tiempos empleados para el entrenamiento de cada arquitectura	55
Figura 16.	Resultado de los tiempos empleados al simular la red.	55
Figura 17.	Resultados del error por cada algoritmo de entrenamiento	57
Figura 18.	Señal original correspondiente a la frase "Aló, Buenos días"	58
Figura 19.	Señal predicha por RNA correspondiente a la frase "Aló, Buenos días"	58
Figura 21.	Resultados al evaluar los parámetros matemáticos generales.	59
Figura 22.	Representación en frecuencia de la señal original	62
Figura 23.	Análisis de frecuencia de la señal predicha	62
Figura 24.	Análisis del error al calcular las frecuencias por ventana	63
Figura 25.	Modelo del sistema de cancelación	66
Figura 26.	Sonómetro Digital Sound Pro	68
Figura 27.	Variación de los parlantes	69
Figura 28.	Configuración final de los parlantes (3 cm de separación)	71
Figura 29.	Diseño final del modelo de cancelación.	71
Figura 30.	Modelo de la superficie para la toma de presión acústica	72
Figura 31.	Experimento en el plano de 132cm <sup>2</sup> .	73
Figura 32.	Experimento en el plano de 200 cm <sup>2</sup>	73
Figura 33.	Escenarios de las pruebas	74
Figura 34.	Potencia de la señal sin cancelar a volumen habitual. Escenario CENTIC	76
Figura 35.	Potencia de la señal cancelada. Escenario 1	78
Figura 36.	Potencia Disipada Plano (0 -200cm) Escenario CENTIC ( 45dB ruido ambiental)	80
Figura 37.	Potencia de la señal a volumen habitual. Escenario 2	82
Figura 38.	Potencia de la señal cancelada. Escenario 2	82
Figura 39.	Potencia Disipada Plano (0 - 132 cm) Escenario CTI (33 dB ruido ambiental)	83
Figura 40.	Parlantes totalmente de frente con una separación de 3cm	86
Figura 41.	Potencia de la señal cancelada predicha con LPC. Escenario 2	90
Figura 42.	Potencia de la señal cancelada predicha con RNA. Escenario 2	90
Figura 43.	Atenuación de potencia, z=0	93
Figura 44.	Atenuación de potencia, z=25	93
Figura 45.	Atenuación de potencia, z=50	94
Figura 46.	Casos de uso sistema total.	97
Figura 47.	Casos de uso interfaz Menú principal.	98
Figura 48.	Ventana del Menú Principal	98
Figura 49.	Ventana de análisis de Parámetros	99
Figura 50.	Casos de uso interfaz Análisis de Parámetros.	100
Figura 51.	Ventana de Análisis de resultados	101
Figura 52.	Ventana de análisis de resultados para un segmento de la señal	102
Figura 53.	Casos de uso interfaz Análisis de Resultados	103
Figura 54.	Interfaz de Resultados de Cancelación	104
Figura 55.	Comportamiento de la cancelación de señales de voz en el entorno	105
Figura 56.	Casos de uso interfaz Resultados de Cancelación	106

## ANEXOS

ANEXO A. Pruebas de algoritmos de entrenamiento	114
ANEXO B. Pruebas de cancelación escenarios 1 y 2	116
ANEXO C. Potencias de cancelación CTI sin ruido ambiental	118
ANEXO D. Pruebas cualitativas con usuario final	119
ANEXO E. Cancelación de señales con técnicas de predicción	123
ANEXO F. Potencia de cancelación de señales con técnicas sin ruido ambiental	126
ANEXO G. Trabajo futuro: DSP	129
ANEXO H. Artículo de predicción	131

## RESUMEN

**TITULO:** SISTEMA DE CANCELACIÓN DE SEÑALES DE AUDIO EMITIDAS POR UN INDIVIDUO EN SUS AREAS CIRCUNDANTES\*

**AUTORES:** ADRIANA ROCÍO GARCÍA FANDIÑO, JULY PAOLA LICONA BLANCO\*\*

**PALABRAS CLAVES:** Coeficientes Predicción Lineal, Redes Neuronales, Reconstrucción, Señales Digitales, Predicción de Voz, Cancelación de Señales, Presión Acústica.

### DESCRIPCIÓN:

En el presente proyecto se permiten estudiar tres etapas importantes para el desarrollo del sistema: la primera etapa consiste en la captura de la señal de audio, la segunda etapa es la predicción de las muestras de la señal de voz para tener lista la señal contraria, necesaria en la última etapa que consiste en realizar la cancelación sumando la señal original con la señal predicha.

La predicción de las muestras de la señal de voz se realiza por dos métodos: Predicción lineal de voz mediante LPC, que por medio de sumatorias y combinación lineal de muestras anteriores se obtienen los siguientes valores y Redes Neuronales Artificiales (RNA) en donde, se entrena una red neuronal capaz de predecir las muestras siguientes a partir de un conjunto de muestras anteriores.

A partir de los resultados de predicción es posible obtener valores matemáticos comparativos entre las dos técnicas tales como: tiempo de respuesta, calidad de la señal predicha, análisis de frecuencia de la voz y características propias de la voz.

La cancelación de la señal permite el estudio de la potencia empleando las resultantes de la predicción, es decir, cancelación cuando se invierte la señal predicha (por LPC ó por RNA), de manera que se puede estudiar el impacto de poner en funcionamiento el sistema de cancelación contrastando los resultados con la cancelación óptima (invertir la señal original y sumar),

---

\* Trabajo de investigación.

\*\* Facultad de Ingenierías Físico Mecánicas. Escuela de Ingeniería de Sistemas e Informática.  
Director: Henry Arguello Fuentes.

## ABSTRACT

**TITLE:** SIGNAL SYSTEM FOR CANCELLATION OF AUDIO DELIVERED BY AN INDIVIDUAL IN ITS SURROUNDING AREAS<sup>\*</sup>

**AUTHORS:** ADRIANA ROCÍO GARCÍA FANDIÑO, JULY PAOLA LICONA BLANCO<sup>\*\*</sup>

**KEY WORDS:** Linear Prediction Coding, Artificial Neural Network, reconstruction, digital signals, Prediction Voice, cancellation of signals, acoustic pressure,

### Description:

This project allowed to study three important stages of the system development: the first stage is the capture of the audio signal, the second stage is the prediction for the samples of the voice in order to get the inverse signal, which is needed for the third and last stage that is the cancellation of the voice, adding the original signal to the signal predicted.

The prediction of the voice signals samples is done applying two methods:

- Lineal Prediction of The Voice using LPC, so the subsequent signal values are obtained making summations and lineal combinations of previous samples.
- Artificial Neural Networks (ARN), which are trained until they are capable of predicting the next samples, given a set of previous samples.

From the result of the predictions it is possible to get mathematical comparative values between the two techniques such as: response time, quality of the predicted signal, frequency analysis of signal and voice characteristics.

The signal cancellation allows the study of the voice power using the resultants of the prediction, i.e., the cancellation obtained when the predicted signal is inverted (using LPC or ARN), so that the impact of having the system running can be analyzed versus the results of the optimal cancellation (obtained by adding the original signal to its inverse).

---

<sup>\*</sup> Trabajo de investigación

<sup>\*\*</sup> Facultad de Ingenierías Físico Mecánicas. Escuela de Ingeniería de Sistemas e Informática.  
Director: Henry Arguello Fuentes.

## INTRODUCCIÓN

Frente al creciente avance de la tecnología en las últimas décadas y con él, el desarrollo de la telefonía celular se ha abierto un campo grande en la investigación debido a que esto trae progreso en las comunicaciones y a su vez problemas anexos que requieren de soluciones prácticas e innovadoras.

Son más de 3 mil millones de persona que utilizan móviles en el mundo\* y un gran porcentaje de ellos no pueden estar sin su localizador en ningún momento del día, se ha convertido en una herramienta imprescindible para todos.

El uso de celulares es algo que está ligado a la mitad de la población y el número de usuarios crece exponencialmente. La contaminación auditiva causada por estos aparatos es algo que afecta a todos por que en todo momento están presentes los usuarios de este medio de comunicación.

En la universidad, en la calle e incluso en la casa siempre hay alguien que utiliza el teléfono celular llegando al punto de ser molesto para las personas circundantes y aumentando esta incomodidad según el momento y lugar donde nos encontremos.

Previendo que el número de usuarios va en aumento se hace necesario indagar una forma de mitigar la incidencia del uso de este medio de comunicación. Se busca solucionar dos problemas estrechamente ligados referentes al creciente uso de teléfonos celulares:

---

\* Encuesta elaborada por la consultora The Mobile World

- Contaminación auditiva
- Privacidad

Por tanto, las investigaciones y trabajos realizados en el alma mater de la universidad deben estar enfocados en resolver problemas que afecten a la sociedad, desde la aplicabilidad de la ingeniería de sistemas en las señales y teniendo lo anterior en mente surge el siguiente interrogante:

¿Cómo aumentar la privacidad y eliminar el ruido y la contaminación auditiva causada por el incrementado uso de los celulares en lugares públicos, donde hay mucha gente o donde es necesario el silencio?

Para resolver este interrogante, primero se debe realizar una investigación sobre las tecnologías y métodos existentes que abordan esta problemática para después proponer una solución que por medio de la academia contribuya a resolver el problema que se plantea.

El trabajo propuesto es una herramienta base que consta de la implementación de técnicas matemáticas útiles para la predicción de señales de voz para utilizarse en un sistema de cancelación del discurso activo en áreas circundantes. El desarrollo del proyecto trae ligada la aplicación de conceptos aprendidos desde el ciclo básico hasta las materias enfocadas en la carrera, de manera que se logran integrar diferentes ramas del programa de estudio para obtener un producto de ingeniería útil a la investigación.

Al final se obtiene la implementación de dos técnicas de predicción con parámetros de medida comparativa, un algoritmo de cancelación de voz con resultados palpables para confirmar que esta metodología contribuye en parte con

el problema planteado, este aporte favorece a que la investigación continúe y se logre a largo plazo que la idea inicial sea un hecho y utilizada cotidianamente.

# 1. DEFINICIÓN DEL PROBLEMA

## 1.1 ANTECEDENTES

La tecnología evoluciona con gran velocidad, el mercado requiere respuestas cada vez más rápidas en un mundo donde los cambios resultan impredecibles, a medida que aumenta la tecnología van con ella en ascenso problemas que afectan al medio ambiente y por tanto al ser humano creador y usuario de estas tecnologías.

Una de las tecnologías que ha tenido un gran desarrollo ha sido la telefonía celular. Desde sus inicios a finales de los 70<sup>\*</sup> ha revolucionado enormemente las actividades que se realizan diariamente. Los teléfonos celulares se han convertido en una herramienta primordial para la gente común y de negocios.

Cuando se desarrolló el primer móvil, la principal premisa era que se pudiera llevar de un lado para otro, hoy en día se ven teléfonos celulares en todos los lugares y situaciones.

La evolución del celular ha permitido disminuir su tamaño y peso, gracias a desarrollos de la electrónica que han logrado hacer baterías más pequeñas y de mayor duración, pantallas más nítidas y de colores, así como el integrar software más amigable.

Debido a todos estos avances tecnológicos la investigación en el campo de la ingeniería de sistemas contribuye notablemente debido a que cada vez se requieren más aplicaciones (software), mejorar las condiciones de los

---

• MARTINEZ Evelio. Telefonía Celular: 15 años de historia. México. En: Revista RED, ISSN 16659597. (febrero 2005); p. .30-32

dispositivos tales como calidad de la señal, tiempo de respuesta, calidad de video, fotografía, así como también se ha mejorado en la comodidad del usuario de este dispositivo, creándose mas necesidades en torno a este invento.

Sin embargo, no todos los efectos de estos dispositivos móviles son positivos existen consecuencias secundarias que afectan a los usuarios activos y pasivos de esta tecnología, tales como la contaminación auditiva y con ella la perturbación en el medio de comunicación.

Desde hace años el ruido se ha convertido en un factor contaminante constante en la mayoría de las ciudades, suponiendo en la actualidad un grave problema con efectos fisiológicos, psicológicos, económicos y sociales.

El principal causante de la contaminación acústica es la actividad humana. El ruido ha existido desde la antigüedad, pero es a partir del siglo pasado, como consecuencia de la Revolución Industrial y tecnológica, del desarrollo de nuevos medios de transporte y comunicación y del crecimiento de las ciudades, cuando comienza a aparecer el problema de la contaminación acústica urbana.

El ruido se define como cualquier sonido calificado, por quien lo sufre, como algo molesto, indeseable e irritante. A su vez, se define la contaminación acústica como aquella que se genera por un sonido no deseado, que afecta negativamente a la calidad de vida y sobre todo, a aquellos individuos que desarrollan actividades industriales y a los que usan con bastante frecuencia determinados dispositivos para poder comunicarse.

## **1.2 FORMULACIÓN**

Es necesario buscar alternativas a los problemas del entorno urbano que progresivamente se hacen más complejos y difíciles de manejar, si la academia por medio de sus investigaciones puede contribuir a disminuir el efecto negativo de los dispositivos de sonido en el entorno, se puede atacar y contribuir en pro de la conservación del medio ambiente.

Como alternativa de solución se diseñó un sistema de cancelación de señales de audio emitidas por un individuo en el área circundante, de forma que sea posible atenuar el ruido en el medio, conociendo el mensaje exclusivamente los directamente interesados.

## **1.3 DESCRIPCIÓN**

El sistema de cancelación diseñado en este proyecto está constituido por tres etapas principales: la primera etapa consiste en la captura de la señal de audio, la segunda etapa es la predicción de las muestras de la señal de voz para tener lista la señal contraria y en la última etapa se realiza la cancelación sumando la señal original con la señal predicha por LPC (Coeficientes de Predicción Lineal) o RNA (Redes Neuronales Artificiales).

En el diseño se hacen pruebas y análisis de dos técnicas de predicción de voz:

- Predicción lineal de voz mediante LPC, por medio de sumatorias y combinación lineal de muestras anteriores se obtienen los siguientes valores.

- Redes Neuronales Artificiales (RNA), con esta técnica se entrena una red neuronal capaz de predecir las muestras siguientes a partir de un conjunto de muestras anteriores.

El objetivo de implantar las dos técnicas de predicción es comparar los parámetros tales como: tiempo de respuesta, calidad de la señal predicha, análisis de frecuencia de la voz, características de la voz, entre otros, con el fin de mirar el desempeño de las técnicas en aplicaciones concretas.

## 2 JUSTIFICACIÓN

Ajustar un sistema de este tipo en diferentes dispositivos de audio resulta de interés para sus portadores, ya que podrían realizar conversaciones o emisiones de señales de audio sin afectar considerablemente a las personas que se encuentren a su alrededor.

El impacto de la propuesta es a nivel de las comunicaciones, en la contribución con la disminución de ruido en el ambiente y en la discreción de las conversaciones de carácter privado, además, se pretende evitar el molesto ruido que causa una persona cuando contesta su celular en medio de una reunión.

Técnicas de este tipo han sido desarrolladas en Japón [10] aunque no han sido exploradas en su totalidad. En este proyecto se desea implementar el sistema con una nueva tecnología que permita obtener resultados similares a los logrados por otras investigaciones.

Este algoritmo se desarrollará basado en técnicas de inteligencia artificial porque por medio de ellas se consiguen resolver problemas relacionados con el reconocimiento de formas o patrones, predicción, codificación, control y optimización entre otras aplicaciones.

### **3 OBJETIVOS**

#### **3.1 OBJETIVO GENERAL**

Implementar una herramienta computacional basada en técnicas de inteligencia artificial para realizar la cancelación de las señales de audio emitidas por un individuo en las áreas circundantes a su posición.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Determinar el estado del arte y las técnicas existentes para realizar la predicción de las muestras de una señal de voz.
- Desarrollar e implementar una técnica de predicción de señales de voz para emplearlo posteriormente en la cancelación del sonido.
- Verificar el funcionamiento de la técnica de predicción seleccionada en un sistema de cancelación de voz.
- Determinar la eficiencia del sistema de cancelación de voz en un ambiente real y controlado.

## 4 MARCO TEÓRICO

### 4.1 COEFICIENTES DE PREDICCIÓN LINEAL LPC [6][13]

#### 4.1.1 Definición

La predicción lineal es una herramienta matemática que permite modelar la voz representando la envolvente espectral digital de la señal de forma comprimida, utilizando la información de un modelo lineal. Es una de las técnicas más potentes del análisis de voz y uno de los métodos más útiles para codificar voz con buena calidad a una tasa de bits relativamente baja.

La importancia de este método radica en la capacidad de proporcionar una aproximación a los parámetros de la voz muy precisas y que es relativamente rápido a nivel computacional.

La idea básica del análisis LPC es que puede hacerse una aproximación simple de la señal de voz como una combinación lineal de muestras pasadas por mínimos cuadrados (definido en un intervalo finito) entre las muestras de la señal actual y las muestras linealmente predichas. Un único coeficiente de predicción puede ser determinado.

El método de predicción lineal proviene de un método de caracterización de parámetros robusto, fiable y exacto, que caracterizan la linealidad del sistema variable en el tiempo.

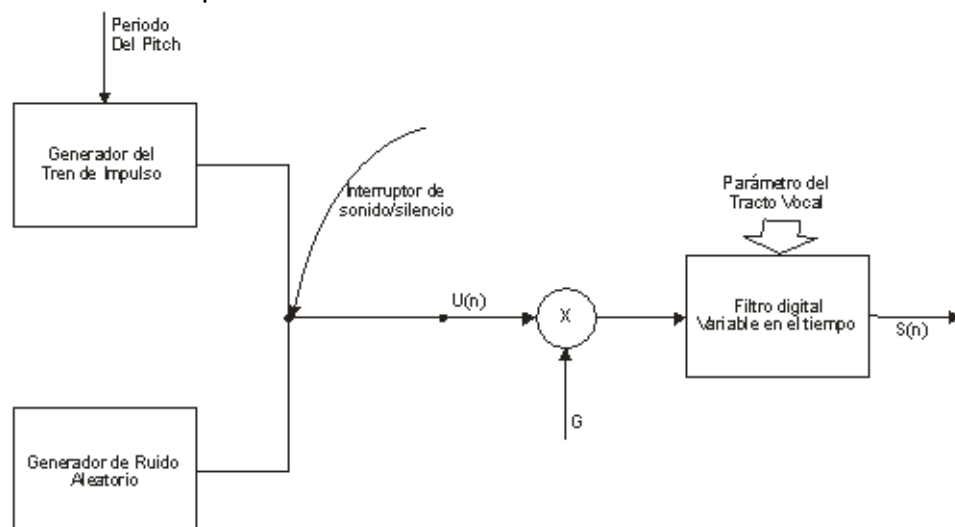
LPC determina los coeficientes de un predictor lineal con principal énfasis en minimizar el error cuadrático. Entre las principales aplicaciones esta el diseño de filtros de codificación y expresión

### 4.1.2 Principios básicos del análisis de predicción

La voz se representa por medio de sistemas lineales. Los modelos suelen formarse utilizando un filtro para separar las partes trascendentales de la señal de voz en cada momento.

El modelo LPC se fundamenta en el modelo de producción de síntesis de voz (Ver figura 1)

Figura 1. Modelo de producción de síntesis de voz



En este diseño el tracto vocal es modelado como la *concatenación de tubos acústicos* de distinto diámetro (con o sin pérdidas).

Esto deriva en un *modelo lineal no estacionario* (ya que las secciones de los tubos van cambiando de acuerdo al fonema que se está emitiendo). El sistema está representado en función de la ecuación 1.

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

Donde,  $G$  es la ganancia de predicción,  $S$  es la señal de voz,  $U$  es la excitación y  $a_k$  son los coeficientes de predicción.

Este sistema es excitado por un tren de impulsos en la señal hablada. Así Los parámetros de este modelo son: sonidos sonoros y no sonoros (clasificados), el periodo fundamental pitch para el discurso, el parámetro de ganancia  $G$  y los coeficientes  $a_k$  del filtro digital. Estos parámetros varían lentamente con el tiempo.

El pitch y la clasificación de sonidos sonoros y no sonoros pueden ser estimados usando uno de muchos métodos basados en análisis de predicción.

Si se utiliza un orden  $p$  adecuado se puede obtener una buena representación de la señal. La mayor ventaja de este modelo es que el parámetro de ganancia  $G$  y los coeficientes  $a_k$  del filtro pueden ser estimados de una manera sencilla y los cálculos computacionales son eficientes por el método de análisis de predicción lineal.

Para el sistema de la figura 1, las muestras de la señal  $s(n)$  están relacionadas con la excitación  $u(n)$  por la simple ecuación de diferencias (ecuación 2)

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2)$$

El método de predicción lineal consiste en predecir una muestra  $n$  de voz a partir de la sumatoria de las  $n-1$  muestras anteriores.

La predicción con los coeficientes  $\alpha_k$  son definidos como un sistema cuya salida es:

$$s[n] = \sum_{k=1}^p a_k s[n-k] \quad (3)$$

El método LPC determina los coeficientes de un predictor lineal con el fin de minimizar el error cuadrático de predicción con el método de mínimos cuadrados. Aplicable en el diseño de filtros y codificación de señales de voz. Se pretende hallar los coeficientes  $a_k$  conociendo el orden  $p$  y la ganancia  $G_u(n)$  de predicción basándose en la ecuación de sumatoria de LPC.

La función del sistema de un predictor lineal de orden  $p$  es el polinomio

$$P[z] = \sum_{k=1}^p \alpha_k z^{-k} \quad (4)$$

El error de predicción  $e(n)$  es definido como:

$$e[n] = s[n] - S[n] = s[n] - \sum_{k=1}^p \alpha_k s[n-k] \quad (5)$$

Se puede observa que el error secuencial de predicción corresponde a la salida del sistema cuya función de transferencia es:

$$A[z] = 1 - \sum_{k=1}^p \alpha_k z^{-k} \quad (6)$$

Comparando las ecuaciones 2 y 5, se puede observar que si la señal de voz obedece el modelo de la ecuación 2 exactamente, y si se hace  $\alpha_k = a_k$  entonces

$s(n) = Gu(n)$ . Por lo tanto el error de predicción del filtro, será inverso al filtro del sistema  $H(z)$  de la ecuación 1.

$$H(z) = \frac{G}{A(z)} \quad (7)$$

El problema básico del análisis de predicción es determinar un conjunto de coeficientes de predicción  $\alpha_k$  directamente de la señal de voz, de tal manera que para obtener una buena estimación del espectro de la señal de voz por el uso de la ecuación 7.

Debido a la variación natural en el tiempo de la señal de voz para predecir los coeficientes se debe estimar a partir de breves segmentos de señal de voz. El enfoque básico es encontrar un conjunto de coeficientes de predicción que logre reducir al mínimo el error cuadrático de predicción en un corto segmento de discurso de forma de onda.

No es inmediatamente evidente que este análisis conduzca a resultados útiles, pero se puede justificar de diversas maneras. En primer lugar, recordar que si  $\alpha_k = a_k$  entonces  $s(n) = Gu(n)$ .

Para las señales de voz esto significa que  $s(n)$  puede consistir en un tren de impulsos, por ejemplo,  $s(n)$  puede ser pequeño la mayor parte del tiempo. Por lo tanto, la búsqueda de unos  $\alpha_k$  que reduzca al mínimo el error, parece coherente con esta observación.

Una segunda motivación para este enfoque, se deriva del hecho de que si una señal es generada por la ecuación 2 con un tiempo invariable y los coeficientes son excitados ya sea por un solo impulso o por un ruido estacionario de entrada,

entonces se puede demostrar que los coeficientes de predicción resultantes de reducir al mínimo el error medio cuadrático (en todo el tiempo) son idénticos a los coeficientes de la ecuación 2.

Una tercera y muy pragmática justificación para el uso del método de mínimos cuadrados como base para estimar los parámetros del modelo es que este enfoque da lugar a un conjunto de ecuaciones lineales que se pueden resolver de una manera eficiente para obtener los parámetros de predicción. Más importante aún el resultado comprende una útil y exacta representación de la señal de voz.

En un tiempo promedio corto el error de predicción está definido como:

$$E_n = \sum_m e_n^2(m) \quad (8)$$

$$E_n = \sum_m (s_n(m) - s'_n(m))^2 \quad (9)$$

$$E_n = \sum_m [s_n(m) - \sum_{k=1}^p a_k s_n(m-k)]^2 \quad (10)$$

Cuando  $s_n(m)$  es un segmento de señal de voz que ha sido seleccionado cercano a  $n$ , por ejemplo:

$s_n(m) = s(m+n)$  El rango de las sumatorias 8 y 10 no está definido en un tiempo infinito, pero el desarrollo se va a realizar en un periodo corto de tiempo, por tanto la suma será siempre igual a lo largo de un intervalo finito.

Además, para obtener un promedio se debe tener en cuenta que hay que dividir la longitud del segmento de la señal. Sin embargo, esta constante es irrelevante para

el conjunto de ecuaciones lineales que se obtendrán posteriormente, por lo cual se omite.

Es posible hallar un valor de  $\alpha_k$  que minimice  $E_n$ , haciendo  $\frac{\partial E_n}{\partial \alpha_t} = 0; t = 1, 2, \dots, p$ , a partir de esto se obtiene la ecuación:

$$\sum_m s_n(m-t)s_n(m) - \sum_{k=1}^p \hat{\alpha}_k \sum_m s_n(m-t)s_n(m-k) \quad 1 \leq t \leq p \quad (11)$$

Donde  $\hat{\alpha}_k$  es el valor de  $\alpha_k$  que minimiza  $E_n$  (Como  $\hat{\alpha}_k$  es unico, se puede remplazar por  $\alpha_k$ ). Si se define:

$$\varphi_n(t, k) = \sum_m s_n(m-t)s_n(m-k) \quad (12)$$

La ecuación 12 puede ser escrita como:

$$\sum_{k=1}^p \alpha_k \varphi_n(t, k) - \varphi_n(t, 0) \quad t = 1, 2, \dots, p \quad (13)$$

Este conjunto de  $p$  ecuaciones y  $n$  incognitas puede resultar eficiente para calcular los coeficientes de predicción lineal  $\{\alpha_k\}$  que minimizan el error cuadrático medio de predicción para el segmento  $s_n(m)$ .

Usando las ecuaciones 10 y 12, el error cuadrático puede ser mostrado como:

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \alpha_k \sum_m s_n(m)s_n(m-k) \quad (14)$$

Y usando 14 se puede expresar  $E_n$  como,

$$E_n = \varphi_n(0,0) - \sum_{k=1}^p \alpha_k \varphi_n(0,k) \quad (15)$$

Por lo tanto, el error total consiste en un error fijo y en un error que depende de los coeficientes de predicción lineal.

Para calcular los coeficientes de predicción óptimos, primero se deben calcular  $\varphi_n(l,k)$  para  $1 \leq l \leq p$  y para  $0 \leq k \leq p$  (16)

Después de realizar esto se procede a solucionar la ecuación 14 para obtener los valores de  $\alpha_k$ . Por lo tanto, el análisis LPC es un método práctico para realizar la predicción de voz, sin embargo, los detalles de los cálculos de los coeficientes y la solución de las ecuaciones puede ser compleja si no se cuenta con un software matemático. Para poder despejar los coeficientes  $\alpha_k$  necesarios en la predicción mediante LPC se utiliza el método de auto correlación\*.

## 4.2 REDES NEURONALES ARTIFICIALES

### 4.2.1 Conceptos [9]

Las RNA corresponden a una forma de computación que trata de resolver problemas que no son solucionables por los métodos tradicionales, problemas que para el ser humano son muy simples, pertenecientes al entorno cotidiano y que las personas resuelven sin darse cuenta (reconocimiento de formas, toma de decisiones, etc. ), dando soluciones robustas y fáciles de interpretar.

---

\* RABINER, Lawrence y SCHAFFER, Ronald. Digital processing of speech signals. New Jersey: Prentice Hall, 1979, p. 401

Cada neurona se comporta como un circuito eléctrico que suma las señales que recibe de otras neuronas y produce una salida de acuerdo con un umbral establecido. Este umbral de activación es determinado por una función de transferencia que debe ser escogido según el problema que se desea resolver.

Estos circuitos están interconectados entre si y estas interconexiones aumentan si la neurona es entrenada. Un caso biológico equivalente es el que se presenta cuando el cerebro identifica un objeto que tiene en frente, cuando los ojos captan un objeto A, por ejemplo, algunos de los sensores de la visión se activan y envían señales a las neuronas ocultas. Las neuronas que se disparan con la señal de entrada aumentan el grado de conexión de ellas. Si el mismo objeto A se presenta una y otra vez, la interconexión de neuronas se refuerza y, por lo tanto, el conocimiento del objeto.

Si se le presenta a la red anterior el objeto A modificado la unión de las neuronas para el conocimiento de tal objeto, es débil. Las neuronas deben entrenarse para reconocer el objeto A en esta nueva presentación. Luego de algunas sesiones de entrenamiento, el sistema neuronal es capaz de reconocer el objeto A en todas sus formas. Si el objeto cambia nuevamente la red de neuronas y el conocimiento se actualizan.

De esta forma se tiene una red artificial capaz de generalizar, es decir, de dar respuesta a cosas que no han sido mostradas o enseñadas directamente. Esta capacidad depende de las entradas empleadas para entrenar la red. Además, para saber si la red en verdad está aprendiendo es necesario estar en constante evaluación. A medida que se va entrenando la red, se realizan pequeños exámenes para verificar que las respuestas obtenidas sean las deseadas.

La RNA recibe valores de entradas y de salida (dependiendo del tipo de entrenamiento) que son los estímulos que le permiten modificar sus conexiones para dar la respuesta correcta. Esta es la gran diferencia que tiene con los sistemas tradicionales. En los sistemas tradicionales algorítmicos, están programados haciéndoles seguir una secuencia de instrucciones predefinida. Las RNA generan ellas mismas sus propias "reglas", para asociar la respuesta a su entrada; es decir, como los humanos aprenden por ejemplos y de sus propios errores.

El conocimiento de una RNA se encuentra en la función de activación utilizada y en los valores de sus interconexiones, comúnmente llamadas pesos sinápticos equivalentes a las conexiones sinápticas del cerebro.

#### 4.2.2 Modelo Neuronal [4]

Para poder emular las redes neuronales biológicas las RNA siguen un modelo matemático. Este modelo es una generalización del propuesto por Mc Culloch y Pitts en 1943:

$$y = f(u(x, w)) = f\left(\sum_{i=1}^N w_i x[n-i]\right) - b \quad (17)$$

La neurona recibe del exterior un umbral y N entradas a las que asocia un conjunto de pesos  $w_i$   $i=1, \dots, N$ . Obtiene la salida aplicando el producto de los pesos por las entradas respectivas más el umbral (o desplazamiento) a una función de activación  $f(u)$ . A continuación se muestra las funciones de activación más comunes:

$$\text{Umbral: } f(u) = \begin{cases} 0 & u < 0 \\ 1 & u \geq 0 \end{cases} \quad (18)$$

$$\text{Lineal a tramos: } f(u) = \begin{cases} 0 & u \leq -\frac{1}{2} \\ u & -\frac{1}{2} < u < \frac{1}{2} \\ 1 & u \geq \frac{1}{2} \end{cases} \quad (19)$$

$$\text{Sigmoidea } f(u) = \begin{cases} 0 & u < 0 \\ 1 & u > 0 \end{cases} \quad (20)$$

$$\text{Gausiana: } f(u) = e^{-\frac{u^2}{\sigma^2}} \quad (21)$$

$u(x,w)$  Recibe el nombre de función base, donde los vectores  $x,w$ :

$$x = \{(x[n-1], \dots, x[n-N])\} \quad w = (w_1, \dots, w_N)$$

Cuando la combinación de las entradas es de tipo lineal,  $u(x,w)$  es una función de base lineal, pero son posibles otras combinaciones, entre las que destacan las funciones de base radial, consistentes en realizar una operación no lineal sobre las entradas, con simetría radial.

### 4.3 ONDAS SONORAS [2]

Una onda sonora es una variación local de la densidad o presión de un medio continuo, que se transmite de unas partes a otras del medio en forma de onda longitudinal periódica o casi periódica.

Las ondas sonoras tienen ciertas propiedades que le permiten rebotar ante una barrera, cambiar de dirección cuando pasan de un medio a otro, sumar sus efectos de una forma muy especial y poder salvar obstáculos o bordear las esquinas de un objeto. Se va a enfatizar en la propiedad de interferencia de ondas por que este concepto se toma como punto de partida para el montaje del sistema de cancelación.

#### **4.3.1 Interferencia de ondas sonoras**

El principio de superposición de ondas expresa que: “Si dos o más ondas se solapan en la misma región del espacio se produce una interferencia y se obtiene una onda resultante cuyas características dependen del desfase de las ondas incidentes”.

Cuando las ondas interfieren entre sí, la amplitud (intensidad o tamaño) de la onda resultante depende de las frecuencias, fases relativas (posiciones relativas de crestas y valles) y amplitudes de las ondas iniciales.

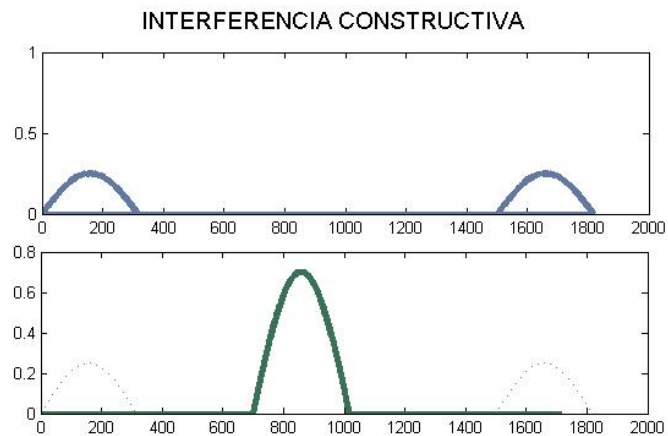
La superposición puede dar lugar a una intensificación o debilitamiento de la onda resultante respecto de las ondas componentes. Para que en la superposición de ondas se produzca interferencia apreciable es necesario que las ondas sean armónicas, es decir que estén en la misma frecuencia. La forma de onda resultante de la superposición de ondas se obtiene sumando algebraicamente cada una de las ondas que componen el movimiento.

##### **4.3.1.1 Interferencia constructiva**

Se produce en los puntos en que dos ondas de la misma frecuencia que se solapan o entrecruzan están en fase; es decir, cuando las crestas y los valles de

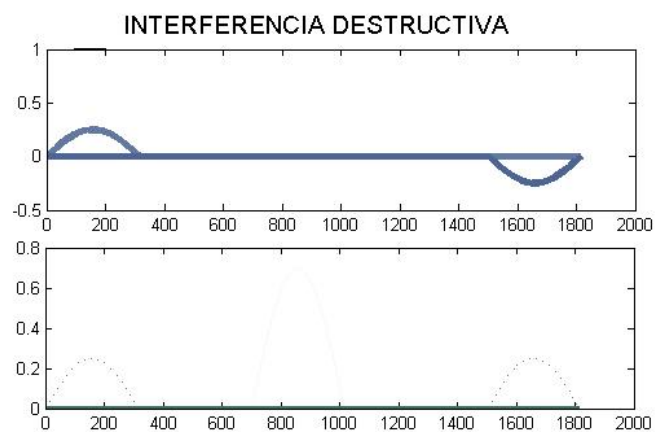
ambas ondas coinciden. En ese caso, las dos ondas se refuerzan mutuamente y forman una onda cuya amplitud es igual a la suma de las amplitudes individuales de las ondas originales como se puede observar en la figura 2

Figura 2. Interferencia constructiva de dos pulsos



#### 4.3.1.2 Interferencia destructiva

Figura 3. Interferencia destructiva de dos pulsos



Se produce cuando dos ondas de la misma frecuencia están completamente desfasadas una respecto a la otra; es decir, cuando la cresta de una onda coincide

con el valle de otra. En este caso, las dos ondas se cancelan mutuamente (ver figura 3)

Cuando las ondas que se cruzan o solapan tienen frecuencias diferentes o no están exactamente en fase ni desfasadas, el esquema de interferencia puede ser más complejo.

Si se superponen ondas sinusoidales de igual frecuencia (pero distinta amplitud y fase) se obtiene una sinusoidal de igual frecuencia pero diferente amplitud y fase. Eventualmente ambas ondas podrían cancelarse, si tuvieran igual amplitud pero a contra fase ( $180^\circ$ ).

La superposición de sinusoidales es de especial relevancia ya que la teoría de Fourier establece que un sonido periódico complejo puede descomponerse como suma de sinusoidales, lo cual implica que las ondas sonoras cumplen esta propiedad.

En general, la interferencia puede producirse con toda clase de ondas. Las ondas de radio interfieren entre sí cuando rebotan en los edificios de las ciudades, con lo que la señal se distorsiona. Cuando se construye una sala de conciertos hay que tener en cuenta la interferencia entre ondas de sonido, para que una interferencia destructiva no haga que en algunas zonas de la sala no puedan oírse los sonidos emitidos desde el escenario.

#### **4.3.2 Frecuencia**

Se refiere al número de ondas que pasan por un punto por unidad de tiempo, para el análisis del sonido, la frecuencia es percibida como el *tono*. Sin embargo, el tono es una magnitud subjetiva y se refiere a la altura o gravedad de un sonido

mientras que la frecuencia es una magnitud objetiva y mensurable referida a formas de onda periódicas. El tono de un sonido aumenta con la frecuencia, pero no en la misma medida. Las frecuencias elevadas se perciben como tonos agudos, las más bajas como tonos graves.

La frecuencia  $F$  de un sonido y su longitud de onda  $L$  están directamente relacionadas con la velocidad de propagación  $V$  en la forma:

$$L = V/F \Leftrightarrow F = V/L \quad (22)$$

La respuesta en frecuencias es un medio importante para determinar la calidad del sonido. La mayoría de los oídos humanos pueden oír frecuencia (sonido) alrededor de 20 HZ a unos 20K HZ. La mayoría de las personas consideran frecuencias graves, las frecuencias por debajo de 310 HZ Frecuencias de rango medio son las frecuencias que van desde unos 310 HZ a unos 12K HZ. Estas frecuencias incluyen el diálogo y la mayor parte de la voz humana, piano, guitarra y otros instrumentos. Frecuencias altas se encuentran en la parte superior de la respuesta de frecuencia del espectro. Por lo general son de aproximadamente 12 K a 20 K o superior. Estos incluyen platillos, notas altas de la voz humana, y algunos instrumentos de cuerda.

### 4.3.3 Amplitud

La amplitud se percibe como *intensidad*, *volumen* o *sonoridad*. Las vibraciones de gran amplitud se perciben como sonidos fuertes, las de pequeña amplitud son sonidos débiles.

La amplitud del sonido tiene una evolución desde el momento inicial hasta que desaparece, y esta evolución concierne tanto a la componente fundamental como a sus armónicos. Su representación es la envolvente del sonido y tiene cuatro partes características: El ataque ("Attack") es la fase en que el sonido va creciendo en intensidad hasta alcanzar un máximo. El sostenido ("Sustain") es una fase en que la intensidad es más o menos constante; en ocasiones después de un pequeño descenso ("Decay") después de la amplitud máxima. Finalmente, la desaparición ("Release") es la fase en que la intensidad va disminuyendo hasta desaparecer.

#### 4.3.4 Potencia sonora

La potencia es la energía dividida por el tiempo (energía en unidad de tiempo). Generalmente expresada en Vatios ("Watts"), en una señal se puede representar como la sumatoria de las amplitudes elevadas al cuadrado

$$P = \sum_{n=1}^n A(x)^2 \quad (23)$$

Donde,  $P$  corresponde a la potencia,  $A$  corresponde a la amplitud de la señal y  $n$  corresponde a la longitud de la señal.

El frente de ondas que se propaga alrededor de un punto transporta energía que es captada por el tímpano, y es la responsable de que se pueda oír. Cuanto más extenso sea este frente (más alejado del punto inicial) la energía estará menos concentrada (el sonido se percibe como más débil). La energía no es muy representativa de lo "Fuerte" que puede resultar un sonido.

La representación más importante es la energía por unidad de superficie. Sin embargo, el oído es un órgano extraordinariamente sensible en este aspecto, pudiendo percibir energías infinitesimales. Se estima que al hablar en tono normal se desarrolla una potencia de  $0.00001\text{ W}$  ( $10^{-5}\text{ W}$ ), y tres veces este valor para un grito.

#### 4.3.5 Intensidad sonora

Esta propiedad, también llamada intensidad acústica, hace que el sonido se capte fuerte o débil. La intensidad acústica es una magnitud que da idea de la cantidad de energía que está fluyendo por el medio como consecuencia de la propagación de la onda.

Se puede expresar matemáticamente como la potencia como unidad de área

$$I = \frac{P}{\text{Área}} \quad (24)$$

Donde,  $I$  es la intensidad y  $P$  se refiere a la potencia

La intensidad  $I$  de un sonido puede medirse mediante la energía que transporta por unidad de superficie (medida en perpendicular a la dirección de propagación); se expresa en  $\text{W}/\text{m}^2$ . Depende de la amplitud y frecuencia de la fuente sonora, y es independiente de cualquier consideración subjetiva (del observador).

Se conoce que la intensidad del sonido más tenue perceptible por el oído humano normal es de  $10^{-12}\text{ W}/\text{m}^2$ , y que  $1\text{ W}/\text{m}^2$  es el umbral del dolor. En base a estos datos, puede afirmarse que el oído humano es realmente notable y que su rango de percepción es increíble. Por ejemplo, suponiendo que la energía se distribuye

uniformemente alrededor del que habla, un interlocutor a **1 m** percibe una intensidad

$$I = \frac{10^{-5}}{4\pi \cdot 1^2} = 79.577 \cdot 10^{-8} \text{ W/m}^2 \quad (25)$$

#### 4.3.6 Nivel de presión acústica

Otra forma de referirse a la energía transportada por un sonido es su nivel de presión SPL ("Sound Pressure Level") o volumen acústico, que depende de la amplitud de la vibración. Es importante significar que la misma frecuencia parece de tono distinto cuando cambia su intensidad y que niveles elevados se consideran dañinos para la salud y pueden llegar a producir sensaciones dolorosas.

El volumen acústico se mide en Bels, abreviadamente B, el volumen acústico  $\beta$  de un sonido de intensidad  $I$  expresado en Bels se define como:

$$\beta = \log \frac{I}{I_0} \text{ (Bels)} \quad (26)$$

Como la unidad resulta grande, se utiliza el decibelio (décima parte del Bel)  $[dB]$  que ha quedado como unidad "de facto" para la medida del volumen acústico.

El volumen acústico  $\beta$  de un sonido de intensidad  $I$  expresado en decibels se define como:

$$\beta = \log \frac{I}{I_0} \text{ (Bels)} \quad (27)$$

El valor de referencia  $I_0$  es  $10^{-12} \text{ W/m}^2$  es la intensidad del sonido más tenue perceptible por el oído humano, que se considera como punto origen para las medidas acústicas. De la definición se deduce que el volumen acústico correspondiente  $I_0$  es:

$$\beta = 10 * \log \frac{I}{I_0} = 10 * \log 1 = 0 \quad (28)$$

La intensidad de  $1 \text{ W/m}^2$  (umbral del dolor), equivalente a una presión sonora de  $120 \text{ dB}$ , y  $1 \text{ dB}$  es la mínima variación de intensidad de un sonido que puede detectar el oído. De la aplicación de la fórmula se deduce que duplicar la intensidad de un sonido (pasar de un valor  $I_1$  a otro  $2 \cdot I_1$ ) supone una variación de  $10 \cdot \log 2 = 3.01 \text{ dB}$ .

La intensidad  $I$  de un sonido es una magnitud absoluta, independiente del observador, mientras que el nivel de presión sonora SPL, es un valor relativo que toma como referencia un nivel de señal relacionado con un observador humano estándar.

La ventaja de utilizar la presión acústica SPL en  $\text{dB}$  frente a la intensidad, es que al estar referida al nivel de percepción humano, es más significativa para hacerse una idea de cuan fuerte es percibido un sonido. Además, tratándose de sonidos para "consumo" humano, voz, música, señales acústicas, etc. el rango a manejar es de 120 dB en vez de 1.000.000.000.000. El nivel de percepción humano para algunos sonidos cotidianos se ven en la tabla 1. Estos valores sirven de parámetro de medidas de sonidos en el ambiente.

Tabla 1. Niveles sonoros y respuestas humanas

<b>sonidos característico</b>	<b>Nivel d presión sonora[dB]</b>	<b>Efecto</b>
Martillo neumático Concierto de Rock	100	Extremadamente fuerte
Camión pesado (15 m) Tránsito urbano	90	Muy molesto-Daño auditivo (8Hrs)
Reloj Despertador (0,5 m) Secador de cabello	80	Molesto
Restaurante ruidoso Transito por autopista Oficina de negocio	70	Difícil uso del teléfono
Aire acondicionado Conversación normal	60	Intrusivo
Tránsito de vehículos livianos (30 m)	50	Silencio
Living Dormitorio Oficina tranquila	40	Silencio
Biblioteca Susurro a 5 m	30	Muy silencioso
	20	
	10	Apenas audible
Estudio de radiodifusión	0	Umbral auditivo

Fuente: Adaptada de Noise pollution clearinghouse, disponible: <http://www.nonoise.org>

El ministerio de salud de Colombia\* establece los niveles sonoros máximos permisibles para prevenir y controlar los efectos adversos ocasionados en la población por la emisión de ruido en una clasificación de diferentes zonas como muestra la tabla 2:

\* Resolución 8321 del 4 de agosto de 1983

Tabla 2. Estándares máximos permisibles de niveles de emisión de ruido expresados en decibeles DB(A)

SECTOR	Estándares máximos permisibles de niveles de emisión de ruido en dB(A)	
	Día	Noche
Sector A. Tranquilidad y silencio	55	50
Sector B. Tranquilidad y ruido moderado	65	55
Sector C. Ruido intermedio Restringido	75	75
	70	60
	65	65
	80	75
Sector D. Zona suburbana o rural de tranquilidad y ruido moderado	55	65

Fuente: Ministerio de Salud de Colombia, Decreto 8321 de 1983.

#### 4.4 SENSIBILIDAD HUMANA

El oído humano es un instrumento registrador extremadamente sensible; es capaz de percibir vibraciones del aire del orden de  $10^{-11}$ , aunque evidentemente esta sensibilidad no es la misma para todos los individuos ni para todas las frecuencias. La sensibilidad del oído es extremadamente dependiente de la frecuencia, registrando una sensibilidad máxima entre 3 y 4 **KHz**. Por encima y debajo de estas frecuencias los sonidos se perciben más débilmente.

Según estos resultados resulta evidente que conocer el SPL de un sonido no proporciona suficiente información sobre su sonoridad ("cuanto" de fuerte lo percibimos); habría que indicar su frecuencia y conocer el grado de sensibilidad a esta última.

El término sonoridad se utiliza para describir el nivel de sensación que produce un sonido. La unidad de sonoridad phon, es la que corresponde a un sonido de **1000 Hz** expresado en **dB**.

#### **4.4.1 Relación señal/ruido**

El ruido o contaminación sónica, es un sonido indeseable de cualquier intensidad, que interacciona e interfiere con el ser humano y que siempre produce sensación molesta y desagradable.

La relación señal / ruido es la relación entre la calidad del sonido al ruido. Esta medición se utiliza para medir muchos dispositivos incluidos los receptores, reproductores de CD, reproductores de DVD, etc. Por lo general, cuanto más alto es la cantidad de decibel (dB) mejor es la calidad de sonido. Por ejemplo, una relación señal/ruido de 90 decibelios, o de 100 se considera alta fidelidad.

## **5 ESTADO DEL ARTE**

### **5.1 PREDICCIÓN DE SEÑALES DE VOZ**

La predicción de señales de voz, es un campo ampliamente utilizado en la digitalización, reconocimiento, reducción de datos, envío de señales de voz, y modelado de la voz, sin embargo, la investigación y los estudios de este tema se han detenido debido a que las técnicas existentes muestran resultados satisfactorios para una gran cantidad de aplicaciones en procesos de voz; es preciso retomar el estudio de estas técnicas para abordar nuevas tecnologías y aplicaciones innovadoras. Por este motivo se realizó la investigación de los estudios recientes, en donde, las principales técnicas encontradas en la literatura son:

La técnica de Predicción Lineal (LPC), técnicas con redes neuronales y la combinación de esta última con otras técnicas como lógica difusa, análisis de cepstrum, wavelet, entre otras. A continuación se describen algunas de estas técnicas con sus aplicaciones.

#### **5.1.1 Coeficientes de predicción lineal**

La codificación de predicción lineal tiene sus orígenes hacia los años 50 cuando se da inicio al análisis probabilístico, esta técnica es producto y aplicación de la programación dinámica. Un resumen detallado de la historia de los predictores lineales se encuentra en [3]

Esta técnica tiene buen desempeño en cuanto a los resultados de predicción, modela el tracto vocal de manera lineal haciendo una similitud entre sumatorias y la estructura de la generación de señales de voz, además permite parametrizar

una señal con un número pequeño de patrones y no requiere demasiado tiempo de procesamiento, con lo que se hace de este análisis una herramienta muy utilizada y aplicada a las tecnologías actuales. A pesar de ser antigua y ser lineal, característica poco ventajosa para modelar una señal de voz que es no lineal.

Las aplicaciones encontradas en la literatura sobre LPC están enfocadas en la caracterización de las señales de voz. Para reconocimiento de patrones [20] [7], caracterización de ciertos sonidos, reducción de los datos en la representación de la señal, análisis de ruido en las señales de voz [17], [19] predicción de LPC para un sistema de cancelación del discurso [10], entre otros.

### **5.1.2 Predicción con redes neuronales**

Como un método alternativo desarrollado en los últimos 10 años, se plantea las redes neuronales. La predicción con RNA ha tenido estudios en los cuales se realizaron investigaciones sobre predicción no lineal de voz para mejorar en detalles en la reconstrucción de la señal. Se destacan trabajos en donde se implementaron funciones de base radial, perceptrón multicapa y redes realimentadas, los cuales se realizaron utilizando metodologías de predicción de lazo abierto, de lazo cerrado e híbridas, teniendo en cuenta que estas ideas se explotan dependiendo de la aplicación en la que se implementen las redes neuronales.

Los estudios encontrados recopilan en su gran mayoría predicción para reconocimiento de patrones, por tal motivo la investigación recae en tener la mejor caracterización de la señal de la voz, es importante resaltar el estudio realizado en la Universidad Politécnica de Cataluña en el año 1998, donde se hizo un completo análisis de la predicción con Redes Neuronales[4].

Para la representación de las características de la señal de voz se han venido realizando numerosos estudios que incluyen la red neuronal multicapa [5] y la red neuronal de base radial [8], el estudio que se hace está enfocado en lograr un óptimo funcionamiento modificando la estructura de la red y los parámetros de entrenamiento para mejorar los rendimientos en comparación a los métodos lineales logrando resultados que demuestran que los procedimientos no lineales tienen mejores prestaciones.

Acerca de los métodos que combinan redes neuronales con otras técnicas se destacan la red neuronal- wavelet de realimentación [18] que tiene la capacidad de aproximar el aprendizaje a largo plazo de señales de voz, tiene mejor funcionamiento que las funciones de base radial, la convergencia es más rápida que las otras redes antes citadas, su estructura es más sencilla aunque el procesamiento de los datos toma mayor esfuerzo. Combinando wavelet con estas redes se mejora la predicción a largo plazo. Tiene características de memoria interna para regeneración y localización de frecuencia en el tiempo. Se pueden obtener resultados óptimos aun si el orden de predicción es bajo y finalmente predice señales inestables y periódicas.

La resultante de unir redes con lógica difusa [16] propone una segmentación periódica con redes neuronales para predecir la voz, son una serie de módulos interconectados de forma encadenada donde cada modulo es ejecutado por pequeñas escalas recurrentes con dinámica interna. Con este trabajo se pretende que la predicción sea dinámica y al ser modular requiera menos recursos. También se obtienen resultados adecuados y sería una buena opción para ser aplicada en proyectos futuros.

Se han realizado numerosos trabajos comparativos entre la predicción lineal con el método no lineal, concluyendo que el método lineal es comúnmente utilizado por

su simplicidad, sacrificando calidad y logrando una predicción aceptable. Por su parte el método no lineal corresponde a sistemas más complejos y exige por tanto mayor esfuerzo computacional, a cambio de esto se esperan resultados más acertados.

Los trabajos encontrados en la literatura han demostrado la superioridad del predictor no lineal que puede ser aplicado con Perceptrón multicapa (MLP), redes de base radial (RBF) o híbridos entre redes neuronales y técnicas como Wavelet o fuzzy.

## **5.2 CANCELACIÓN DE SEÑALES**

La investigación principal en este tema recae en cancelación de ruido exterior, numerosos estudios se han realizado para evitar el eco de las comunicaciones entre locutores y receptores, para que las señales de voz lleguen de la manera más clara posible sin pérdida de información y sin distorsión alguna al escucharlas. [11] [12]

En sistemas de reconocimiento es importante cancelar el ruido circundante o información extra que traen consigo las señales de voz; los sistemas de diferentes micrófonos utilizados para las diversas actividades de las telecomunicaciones tienen investigaciones en este tema también, sistemas en los que es necesario incorporar parlantes para llegar a realizar una atenuación del ruido exterior. [1]

En cuanto a la cancelación de discurso activo en el radio circundante en [10] se propone un método de cancelación del discurso invirtiendo una señal predicha, cuya fuente sería un altavoz ruidoso secundario en donde se logra atenuar la emisión de discurso en el exterior, la principal motivación se reduce a sistemas de telefonía celular, dictados o comunicación por medio de micrófonos en donde no

se requiere que el discurso sea irradiado en el espacio circundante, sólo en el micrófono. En el trabajo se demostró que la cancelación es posible y se da con ciertas condiciones: el sistema es dependiente del locutor, de los altavoces y la ubicación del parlante de cancelación debe ser totalmente de frente a la fuente primaria, utilizan predicción mediante LPC para predecir la muestra que podría ir unos milisegundos después, esta señal era creada por separado para después enviarla simultáneamente con la señal emitida por el parlante.

### **5.3 TECNOLOGÍAS QUE ABORDAN LA PROBLEMÁTICA**

La solución al problema recae en generar nuevas tecnologías que permitan la cancelación o atenuación del discurso activo en el área circundante, existen pocos estudios pero se buscan nuevas formas de abordar el problema, además de crear un dispositivo que elimine el ruido emitido mediante cortinas y ondas sonoras que creen interferencia destructiva, se está investigando en la manera como mediante el sistema nervioso es posible enviar señales a las cuerdas vocales sin hacer ruido alguno, las señales se recogen por medio de cables y dispositivos electrónicos y son llevadas a un computador donde se traducen en una voz artificial, de esta manera es posible comunicar un mensaje o contestar una llamada sin emitir ruido alguno en el entorno en donde se encuentre el emisor, hasta el momento el sistema está en desarrollo y se limita a determinado número de frases. Se espera que al finalizar el año el sistema pueda reproducir cualquier frase sin limitaciones [15].

## **6 DESARROLLO DEL PROYECTO**

El proceso a seguir para lograr los objetivos del proyecto consiste en implementar las técnicas estudiadas siguiendo un proceso que abarca la captura, predicción y cancelación de la señal como se ilustra a continuación:

### **6.1 CAPTURA DE LA SEÑAL**

La señal de voz básicamente está constituida por ondas de presión producidas por el aparato fonador humano. Con el fin de capturar este tipo de señal analógica, se utiliza un micrófono que convierte la onda de presión sonora en una señal eléctrica.

Es necesario definir la frecuencia de muestreo y la cuantificación de los bits para tener en cuenta en el momento de manipular las señales de voz, para el presente trabajo las muestras fueron tomadas con una frecuencia de 8000Hz y una cuantificación de 16 bits, este muestreo se realiza debido a que la aplicación a futuro es pensada para dispositivos móviles en donde la calidad del muestreo es baja debido a la compresión que se debe realizar a los datos.

Estas muestras se recopilaron en una base de datos la cual corresponde a frases comúnmente utilizadas en conversaciones telefónicas, las muestras estudiadas pertenecen a individuos diferentes, al realizar el entrenamiento se debe tener en cuenta, que para cada persona las características de voz varían, lo que influye en una buena predicción de la señal.

### 6.1.1 Recursos empleados en la captura

La siguiente lista corresponde a dispositivos utilizados en la captura de señales del proyecto

- *Computador DELL INSPIRON 640m*

Tabla 3. Especificaciones técnicas del computador

<b>Referencia</b>	<i>INSPIRON 640m</i>
<b>Nombre del sistema operativo</b>	Microsoft Windows XP Professional
<b>Versión</b>	5.1.2600 Service Pack 2 Compilación 2600
<b>Tipo de sistema</b>	X86-based PC
<b>Procesador (Centrino duo)</b>	x86 Family 6 Model 14 Stepping 8 Genuine Intel ~1596 MHz
<b>BIOS Versión/Fecha</b>	Dell Inc. A06, 2006/11/17
<b>Memoria física total</b>	1,024.00 MB
<b>Memoria física disponible</b>	305.74 MB
<b>Memoria virtual total</b>	2.00 GB
<b>Memoria virtual disponible</b>	1.91 GB
<b>Espacio de archivo de paginación</b>	2.38 GB

- Componente de sonido

Tabla 4. Descripción de la tarjeta de audio

<b>Nombre</b>	SigmaTel High Definition Audio CODEC
<b>Fabricante</b>	SigmaTel
<b>Controlador</b>	c:\windows\system32\drivers\sthda.sys (5.10.4995.1 nd446 cp1, 1.10 MB (1,156,648 Bytes), 2006/12/12 05:07 p.m.)

- Especificaciones del micrófono para pruebas directas

Tabla 5. Micrófono de escritorio omnidireccional Genius MIC-01<sup>a</sup>

<b>Impedancia máxima de salida</b>	2,2 kW(máx)
<b>Sensibilidad</b>	-62 dB +-4 dB( a 1Khz)
<b>Respuesta de frecuencia</b>	De 100Hz a 10kHz
<b>Relación señal/ ruido</b>	> 40 dB (a 1kHz)

- Especificaciones micrófono del CENTIC

Tabla 6. Micrófono Cobalt CO 11

<b>Referencia</b>	Electro-Voice Cobalt CO 11
<b>Respuesta de frecuencia</b>	50Hz 20kHz / - 3 dB
<b>Sensibilidad</b>	4mV/Pa (-48dBV) re 1 Pa
<b>Dimensiones (alto x diámetro máximo)</b>	6.69 "x 2.09" (170 x 53mm)

- *Editor de señales de voz*

*Cool Edith pro 2.0 a una frecuencia de 8000, tipo mono y 16 bits.*

### 6.1.2 Descripción del ambiente

Para realizar la captura de la señal se hizo necesario tener en cuenta el entorno en el cual se grabaron las voces. Delimitando el proyecto en primera instancia a señales en un ambiente controlado. La toma de las muestras se realizó en el estudio de grabación del CENTIC\* de la Universidad Industrial de Santander

---

\* CENTIC: Centro de Tecnologías de Información y comunicación

(Figura 4). El nivel del ruido ambiental corresponde al aire acondicionado del lugar de grabación

Tabla 7. Características de la sala de grabación del CENTIC

<b>Tamaño:</b>	3,95 m X 2,6 m X 2,70 m
<b>Número de Persona:</b>	1
<b>Potencia del ruido Ambiental:</b>	45 dB
<b>Micrófono:</b>	Cobalt CO 11
<b>Tarjeta de Sonido:</b>	M – Audio Firewire 1814

Figura 4. Sala de Grabación CENTIC



## 6.2 PREDICCIÓN DE LA SEÑAL

Una vez recopiladas las señales de voz, el siguiente paso es desarrollar el algoritmo de predicción. Predecir la señal de Voz consiste en determinar el patrón en el instante  $t+1$  a partir de los patrones en los instantes  $t$ ,  $t-1$ ,  $t-2$ , etc. De esta manera basándose en los primeros datos de la señal se puede predecir la siguiente muestra de voz.

Estudiados los procedimientos de predicción de señales se implementaron dos técnicas encontradas en la literatura paralelamente con el objetivo de determinar cuál de ellas presenta mejores prestaciones, las técnicas empleadas son:

- Predicción de voz mediante coeficientes de predicción lineal (LPC)
- Predicción por redes neuronales artificiales (RNA).

Previo a la predicción de voz se realizó una normalización fija de la señal con el fin de garantizar un acoplamiento y estandarización de los datos, de esta manera estar seguros de los datos de entrada adecuados para la predicción en el instante  $t+1$ . La relación de normalización se muestra a continuación (ecuación 29):

$$S_n = \frac{x(t)}{\max(\text{abs}(x(t)))} \quad (29)$$

Donde  $x(t)$  es la señal original y  $S_n$  corresponde a la señal normalizada,  $\max()$  devuelve el máximo de un vector y  $\text{abs}()$  el valor absoluto del argumento.

Esta normalización permite que la búsqueda de gradiente durante el aprendizaje se realice de manera más efectiva.

## 6.2.1 Predicción de voz mediante coeficientes de predicción lineal

### 6.2.1.1 Procedimiento para la predicción

1. Se determina la frecuencia máxima de la señal, generalmente es de 8Khz y el número de coeficientes  $p$  para hallar por cada ventana, el tamaño de la ventana es de 120 para este estudio.

## 2. Cálculo de los coeficientes de predicción lineal $a_k$

2.1. Se determina el valor de  $M$  (constante que indica cada cuantas muestras se calculan los coeficientes, es el corrimiento del vector  $a_k$ )

2.2. Por cada segmento de la señal se calculan  $p$  valores de  $a_k$ , por el método de auto correlación regresiva para encontrar dichos coeficientes\*.

## 3. Predicción de la señal:

Para predecir  $n$  muestras se utiliza la siguiente sumatoria:

$$S[n] = \sum_{k=1}^p a_k s[n-k] \quad (30)$$

Cuyos parámetros son:  $s[n-1]$ , muestras anteriores del valor siguiente a calcular,  $p$  que representa el tamaño de la ventana, y  $a_k$  son los coeficientes de predicción.

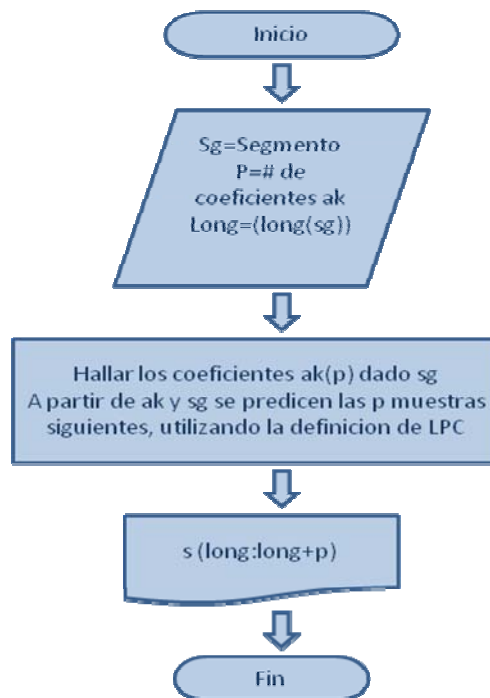
Para predecir el primero y segundo conjunto de tamaño  $M$  de muestras se utiliza  $a_k[1]$ , mientras se llena el búfer de datos que permitan calcular  $a_k[2]$ , que permitirán calcular el tercer grupo de  $M$  muestras a partir de este momento el conjunto  $M$  se halla con  $a_{n-1}[p]$  hasta llegar al tamaño de la señal.

---

\* RABINER, Lawrence y SCHAFFER, Ronald. Digital processing of speech signals. New Jersey: Prentice Hall, 1979, p. 401

Los resultados de la sumatoria son almacenados en un vector que corresponde a la señal predicha. Cada segmento de la señal se predice como se muestra en el diagrama de flujo (Figura 5):

Figura 5. Diagrama de flujo del algoritmo de predicción por LPC



Gráficamente,

Figura 6. Gráfica segmento de entrada y muestras predichas por LPC



### 6.2.1.2 Resultado del método de predicción lineal

Implementado el algoritmo se procedió a probarlo con la variación de los parámetros,  $M$  (número de muestras a hallar) y  $p$  (coeficientes de predicción lineal), obteniendo resultados considerablemente buenos para el objetivo de predicción, la siguiente tabla muestra los datos obtenidos al hacer variaciones de parámetros:

Tabla 8. Pruebas de variación de parámetros con LPC

P	M	Err. V1	T1	Err. V2	T2	Err. V3	T3	Err. V4	T4	Err. V5	T5	Err. Prom.
4	1	2,661E-04	105,04	2,76E-04	230,2062	2,058E-04	317,49	2,37E-04	220,86	2,23E-04	278,71	2,41592E-04
	2	2,610E-04	40,76	2,69E-04	92,814	1,924E-04	131,27	2,27E-04	91,973	2,16E-04	121,02122	2,33286E-04
	3	2,53E-04	28,63	2,59E-04	63,674	1,924E-04	92,03	2,18E-04	67,306	2,07E-04	87,105	2,25714E-04
	4	2,396E-04	23,67	2,52E-04	56,713	1,843E-04	78,78	2,09E-04	54,912	1,96E-04	67,8182	2,16242E-04
8	2	2,532E-04	62,48	2,46E-04	138,5	1,749E-04	202,68	1,70E-04	140,05	2,21E-04	154,74	2,12990E-04
	4	2,293E-04	30,04	2,24E-04	67,94	1,557E-04	93,25	1,56E-04	93,24	1,95E-04	86,362	1,91948E-04
	6	1,981E-04	22,71	2,05E-04	53,74	1,378E-04	69,81	1,38E-04	73,44	1,74E-04	63	1,70532E-04
	8	1,715E-04	20,11	1,84E-04	48,1	1,185E-04	68,68	1,17E-04	44,203	1,42E-04	65,1246	1,46532E-04
16	4	2,241E-04	39,99	2,10E-04	86,36	1,412E-04	131,22	1,41E-04	120,68	2,00E-04	111,904	1,83352E-04
	8	1,633E-04	23,25	1,69E-04	54,99	1,012E-04	75,61	1,10E-04	51,328	1,51E-04	64,232	1,38862E-04
	12	1,250E-04	19,98	1,36E-04	48,34	8,154E-05	66,98	9,28E-05	43,83	1,14E-04	58,38	1,09705E-04
	16	8,463E-05	18,96	1,22E-04	44,25	8,669E-05	59,50	8,67E-05	62,18	1,04E-04	54,45	9,67034E-05
20	5	2,116E-04	32,58	2,02E-04	85,53	1,385E-04	113,61	1,38E-04	77,534	1,89E-04	91,79	1,75700E-04
	10	1,522E-04	22,60	1,47E-04	52,401	9,494E-05	69,28	1,01E-04	46,865	1,33E-04	62,98	1,25610E-04
	15	1,370E-04	19,80	1,25E-04	48,58	9,212E-05	67,28	9,11E-05	40,796	1,08E-04	55,99	1,10528E-04
	20	8,493E-04	18,99	1,05E-04	44,71	8,348E-05	58,88	8,78E-05	38,286	8,20E-05	573	2,41534E-04

En donde:

Err. V1, V2, V3, V4, V5, son los errores cuadráticos medios calculados tomando como parámetro teórico la Voz original, los parámetros  $V_n$  corresponden a diferentes voces con las cuales se mide el funcionamiento del método;  $T_n$ ,  $n=1, 2, 3, 4, 5$  son los tiempos de simulación correspondientes para cada voz, y el error

promedio se calcula como el promedio de los errores cuadráticos medios de las voces así:

$$Prom\ error = \frac{\sum_{n=1}^N E_n}{N} \quad (31)$$

El error cuadrático medio para cada señal, se calcula de acuerdo a la siguiente fórmula:

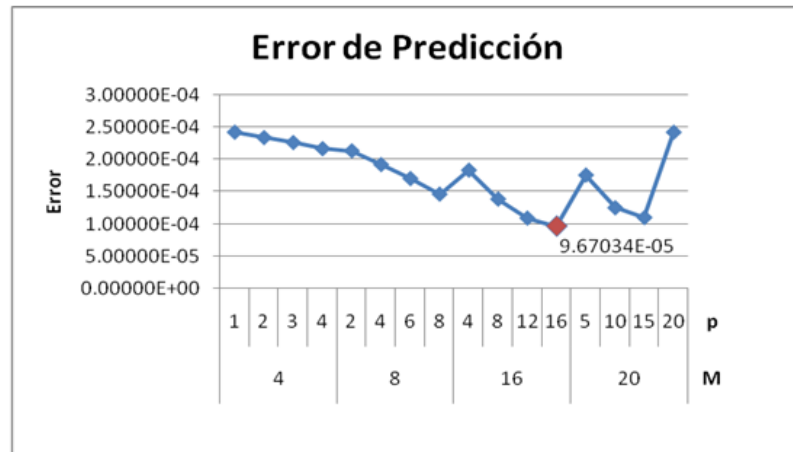
$$E2M = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x})^2}{n}} \quad (32)$$

Donde  $x_i$  corresponde a la señal original y  $\hat{x}$  a la predicha,  $n$  es la longitud de la señal original. Si el valor de  $E2M$  tiene valores próximos a cero, es un buen estimativo porque la señal predicha se acerca a la original.

El método de predicción lineal independiente de la variación de parámetros alcanza errores del orden de  $10^{-4}$ , estimativo aceptablemente bueno debido a que un orden de  $10^{-4}$  es aproximado a cero, un error mínimo que se traduce en resultados satisfactorios, además, el tiempo de simulación es bajo, pero entre más pequeño es el valor de  $M$ , más tiempo de procesamiento es necesario.

A continuación se muestra gráficamente el error (Figura 7), parámetro primordial para escoger una buena configuración de LPC. Variando:  $M$  constante que indica cada cuantas muestras se calculan los coeficientes, es el corrimiento de  $a_k$  y  $p$  que es la constante que indica el número de coeficientes que se van a utilizar para calcular cada muestra de la señal.

Figura 7. Gráfica del error variando p y M para elegir la mejor configuración



Para una configuración de  $p=16$  y  $M=16$  se encuentra el menor error ( $9,97034 \times 10^{-5}$ ), además, se puede observar que los errores de predicción son muy bajos.

Finalmente, se muestra el resultado gráficamente con la frase "Aló, buenos días" de la base de datos.

Figura 8. Señal original correspondiente a la frase “Aló, Buenos días”

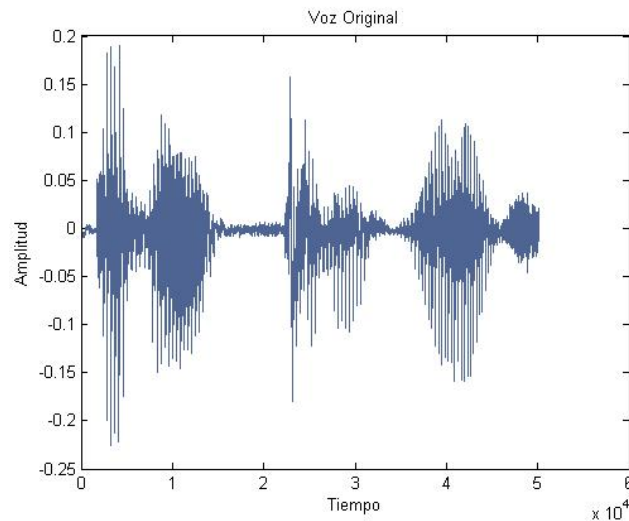
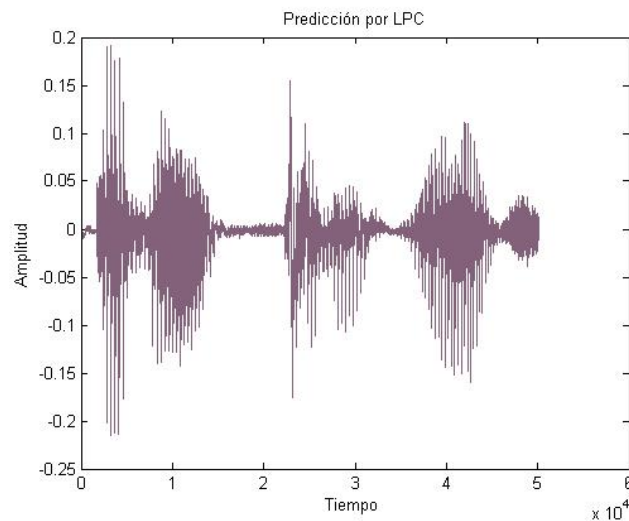


Figura 9. Señal predicha con LPC correspondiente a la frase “Aló, Buenos días”



Analizando las figuras 8 y 9 es notable que la predicción por LPC arroja resultados satisfactorios, gráficamente son dos señales muy similares y al escucharlas el receptor no percibe la diferencia entre una y otra.

Sin embargo, la predicción Lineal de voz solo permite la variación de dos parámetros el orden de predicción  $p$  y la variación  $M$  de muestras, es una técnica

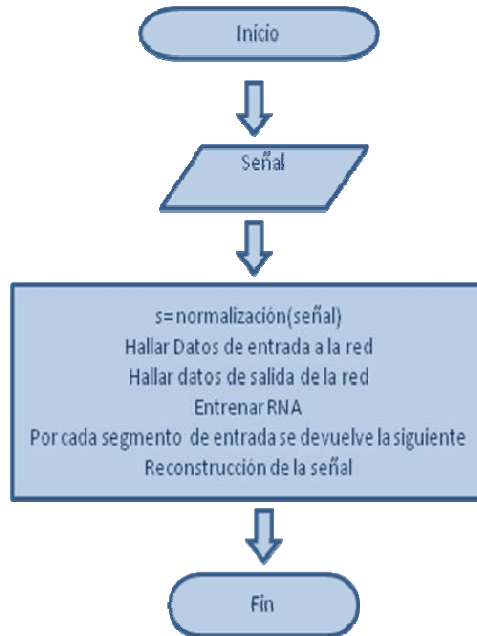
lineal que se reduce a el cálculo de los coeficientes y la reconstrucción de la señal por medio de una sumatoria, no brinda una posibilidad de seguir haciendo estudios y perfeccionándola, aunque ofrece buenos resultados en el momento de obtener la señal predicha en cuanto a precisión y tiempo de simulación.

### **6.2.2 Predicción de la señal de voz mediante Redes Neuronales Artificiales**

Continuando con el estudio de la predicción se procede a implementar la técnica de Redes Neuronales Artificiales, predecir con RNA brinda mayores posibilidades de investigación y de mejoras en la calidad de la técnica, la predicción por medio de RNA consiste en entrenar una red con un conjunto de muestras de determinada señal de manera que retorne el siguiente dato.

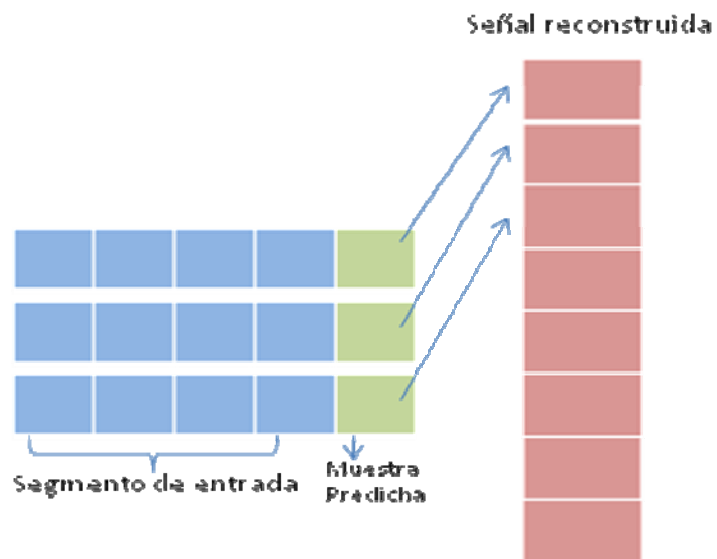
El conjunto de datos de entrada es una matriz de  $20 \times n$  muestras donde  $n$  es la longitud de la señal, lo cual indica que por cada 20 muestras devuelve la muestra 21. Este proceso se realiza con una señal de longitud variable según la frecuencia de la señal pero lo suficientemente larga de manera que la red tenga un conjunto de entrenamiento considerablemente bueno para usar en la predicción. El proceso se muestra en el siguiente diagrama de flujo:

Figura 10. Diagrama de flujo del algoritmo de predicción con RNA



Gráficamente, la señal se predice y reconstruye así:

Figura 11. Gráfica segmento de entrada y muestras predichas por RNA



Con el fin de escoger una arquitectura con resultados óptimos, que logre predecir la señal de voz, se hizo necesario realizar múltiples entrenamientos en los cuales

se probó cada arquitectura repetitivamente buscando los mejores pesos y la mejor configuración, de esta manera se puede garantizar una red con buenas prestaciones relacionadas con el error de predicción (Generalización) y el tiempo de convergencia.

El desarrollo de esta técnica, brinda la posibilidad de realizar pruebas variando diferentes parámetros tales como: número de neuronas, número de capas, función de entrenamiento, número de épocas entre otros.

### 6.2.2.1 Elección de la arquitectura adecuada

Por la dimensión y las características del proyectos se escogió una arquitectura de dos capas ocultas y una de salida, se varió el número de neuronas por cada capa, evaluando la función de rendimiento (FR), Error cuadrático medio de acierto (E2M(a)), Error cuadrático medio de generalización (E2M(g)), el tiempo de entrenamiento (Te) y el tiempo de simulación (Ts), utilizando 2000 épocas para cada entrenamiento.

La notación para la arquitectura es la descrita:  $n_1 \times n_2 \times 1$ ;  $n_1$  equivale al número de neuronas de la primera capa oculta,  $n_2$  número de neuronas de la segunda capa oculta, y 1 es la neurona de la capa de salida.

Los resultados obtenidos se muestran en la Tabla 9:

Tabla 9. Elección de la mejor arquitectura

	Arquitectura	F R	E2M(a)	E2M(g)	Te	Ts
1	5x5x1	0,000294523	5,11E-04	0,000506	114,1632	157,97
2	8x3x1	0,000247768	3,30E-04	0,000925	125,3432	146,37

3	2x2x1	0,000201745	7,55E-04	0,001475	65,095771	158,91
4	2X10X1	0,000333627	3,19E-04	0,0007636	185,31	140,68
5	5x10x1	0,000292734	0,0012	0,001575	180,046	158,9
6	6X7X1	0,000200888	3,45E-04	0,0005873	214,59	143,45
7	4X9x1	0,000230226	2,13E-04	0,0002138	218,9	152,5
8	15x5x1	0,000144523	2,40E-04	0,0005965	375,8268	166,44
9	10x5x1	0,000150803	8,49E-04	0,0004231	243,34	153,63
10	10x10x1	0,000140017	2,38E-04	0,000662	242,043	158,438
11	10x15x1	0,000204027	6,76E-04	0,0011327	391,57	149,91
12	8x5x1	0,000186301	4,41E-04	0,0002085	159,6553	158,66

El error cuadrático medio al igual que en LPC se halla matemáticamente así:

$$E2M = \sqrt{\frac{\sum (x_i - \hat{x})^2}{n}} \quad (33)$$

En donde  $x_i$  corresponde a la señal original y  $\hat{x}$  a la predicha,  $n$  es la longitud de la señal original.

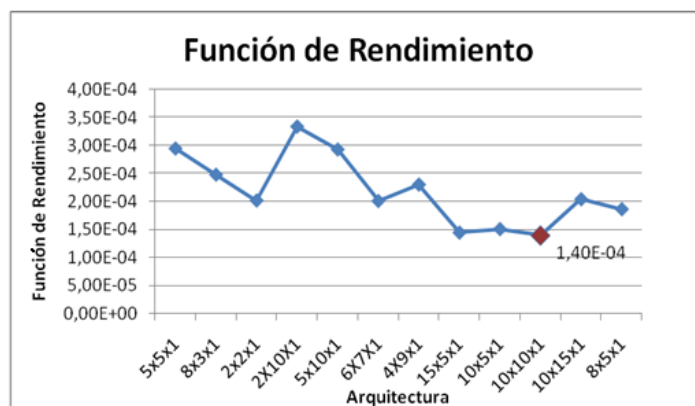
En este caso, el error cuadrático medio de acierto se realiza con respecto a la señal con la que fue entrenada la red, y el error cuadrático medio de generalización con respecto a las señales que no conoce la red, un error pequeño corresponde a una buena generalización o un buen acierto, al igual que en LPC son errores del orden de  $10^{-4}$ , errores mínimos que tienden a cero.

La función de rendimiento (FR) indica un estimativo de la manera en que la red se encuentra aprendiendo, es el error de aprendizaje de la red, si es bajo cercano a cero se estima que la red se encuentra aprendiendo y que posiblemente llegue a generalizar bien.

De los datos tomados en la tabla se grafican los parámetros para escoger la mejor arquitectura:

- *Función de Rendimiento*

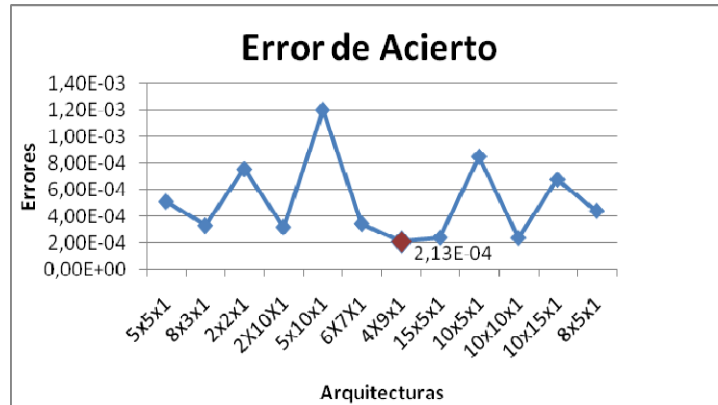
Figura 12. Resultado de la función de rendimiento para las diferentes arquitecturas



Este valor permite tener una idea de qué arquitectura de red puede llegar a ser la más apropiada porque muestra el comportamiento de entrenamiento de la red, teniendo en cuenta sólo esta medida la arquitectura apropiada debería ser 10 neuronas en la primera y segunda capa oculta y 1 neurona en la capa de salida (10x10x1) esta configuración tiene el error más cercano a cero como puede observarse en la figura 12. Sin embargo, el valor de la función de rendimiento para estas arquitecturas son del mismo orden, por tanto no marca una diferencia significativa entre ellas. Además, este parámetro no precisa información sobre si ésta aprende, generaliza o memoriza, por tal motivo es necesario simular la red para apreciar la generalización para la arquitectura escogida.

- *Error de Acierto*

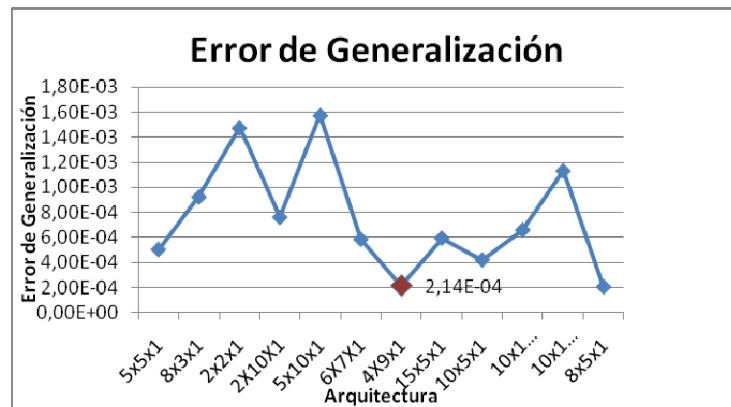
Figura 13. Resultado de los aciertos para las diferentes arquitecturas



El error de acierto (figura 13) muestra el desempeño de la red cuando ella es probada con los datos que se utilizaron para el entrenamiento, esta medida permite verificar en primera instancia el funcionamiento de la red, para esta prueba la mejor arquitectura es 4x9x1, no coincide con la función de rendimiento escogida anteriormente, sin embargo, el valor de la función de rendimiento de esta arquitectura es admitido porque está en el mismo orden que la de menor valor

- *Error de Generalización*

Figura 14. Resultado de la generalización para las diferentes arquitecturas



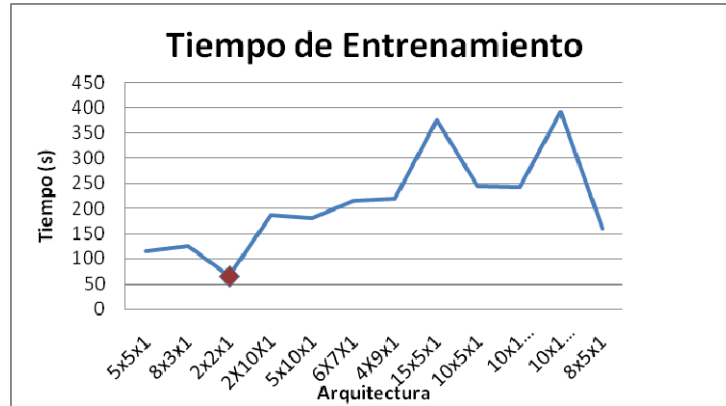
Este es el parámetro más importante (figura 14) debido a que mide la extrapolación del conocimiento adquirido en el entrenamiento de la red, es decir, mide el funcionamiento de la red con muestras diferentes a las de entrenamiento, de esta manera al simularse con una frase diferente se obtienen los resultados adecuados.

Para calcular el grado de generalización de la red se emplearon 9 frases diferentes a la señal de entrenamiento, se estimó un promedio de error después de simular las señales de este conjunto.

- *Tiempo de entrenamiento*

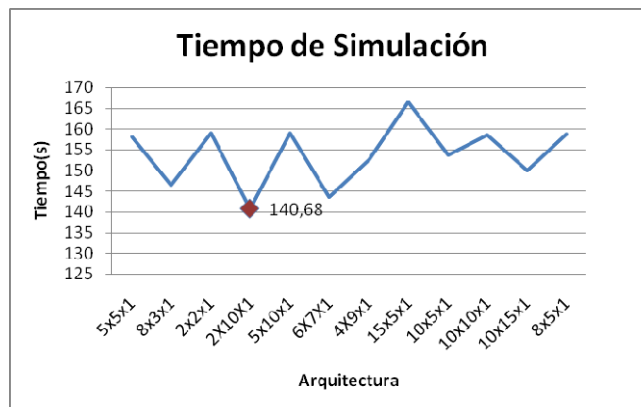
El tiempo de entrenamiento (figura 15) indica los segundos necesarios para el entrenamiento de cada arquitectura, el mejor resultado fue la arquitectura 2X2X1, como se esperaba ya que entre más sencilla es la arquitectura menor es el tiempo de entrenamiento.

Figura 15. Resultado de los tiempos empleados para el entrenamiento de cada arquitectura



- *Tiempo de Simulación*

Figura 16. Resultado de los tiempos empleados al simular la red.



Esta medida (figura 16) indica los segundos que se necesitan en determinada máquina para simular una señal de voz, este tiempo depende tanto de la arquitectura del ordenador donde se realiza la simulación como de la longitud de la

señal que se esté simulando, para este caso se calculó el tiempo de simulación para la expresión “Hola, buenas tardes doctor”, con una frecuencia de 8 KHz y 74743 muestras.

En conclusión, aunque la arquitectura 10x10x1 brinda mejor prestación en cuanto a la función de rendimiento, se escoge la arquitectura 4x9x1 ( 4 neuronas en la primera capa oculta , 9 en la segunda y una de salida) debido a que tiene mejores prestaciones en acierto y generalización, el tiempo de entrenamiento no es un parámetro indispensable para escoger la arquitectura, debido a que una vez entrenada la red solo es necesario simular con los pesos obtenidos. El tiempo de simulación es muy importante para la aplicación que se implementará a futuro, pero en este proyecto no es fundamental porque la arquitectura de la máquina en que se simula no es comparable en rendimiento con el procesador digital de señales (DSP) en el que en próximos trabajos se pretende implementar el desarrollo de esta investigación.

#### **6.2.2.2 Optimización de la red**

Una vez escogida la arquitectura es preciso evaluar si se puede optimizar la red, cambiando parámetros que no se tocaron en las pruebas anteriores tales como: el algoritmo de entrenamiento y el número de épocas.

Se realizaron experimentos en los cuales se manipularon diferentes algoritmos de entrenamiento y un estimativo de épocas para cada uno de ellos, en el ANEXO A se pueden observar los algoritmos con mejores resultados.

A partir de estas pruebas se escogió por cada algoritmo de entrenamiento el mejor número de épocas concluyendo que a partir de determinado número de ellas la función de rendimiento no variaba significativamente.

Teniendo el número de épocas adecuadas para cada algoritmo de entrenamiento se simuló cada uno con 5 frases diferentes para evaluar el error de generalización teniendo un criterio para escoger el más apropiado. Elegir el algoritmo de entrenamiento tiene repercusión en la convergencia rápida y aceptable de la red.

Los resultados de los procedimientos anteriores se muestran en la Tabla 10 y la figura 17.

Tabla 10. Elección del mejor algoritmo de entrenamiento

Algoritmos de entrenamiento	Épocas	Error de Generalización
GDX – <b>Traingdx</b> – Variable Learning Rate Backpropagation	2000	4,37E-04
LM – <b>Trainlm</b> – Levenberg – Marquardt	100	3,31E-05
BFG – <b>Trainbfg</b> – BFGS Quasi – Newton	300	2,34E-04
SCG – <b>Trainscg</b> – Scaled Conjugate Gradient	400	2,50E-04
CGF – <b>Traincgf</b> – Fletcher-Powell Conjugate Gradient	100	3,81E-04
OSS – <b>Trainoss</b> – One-Step secant	500	1,07E-04

Figura 17. Resultados del error por cada algoritmo de entrenamiento



El error de generalización se disminuye en una potencia de diez, con trainlm, algoritmo que utiliza menor número de épocas, lo que repercute en menor costo computacional y menor tiempo de entrenamiento.

En último lugar, para observar el desempeño de la red escogida, se muestran los resultados al simular la frase “aló, buenos días” de la base de datos. (Ver Figuras 18 y 19)

Figura 18. Señal original correspondiente a la frase “Aló, Buenos días”

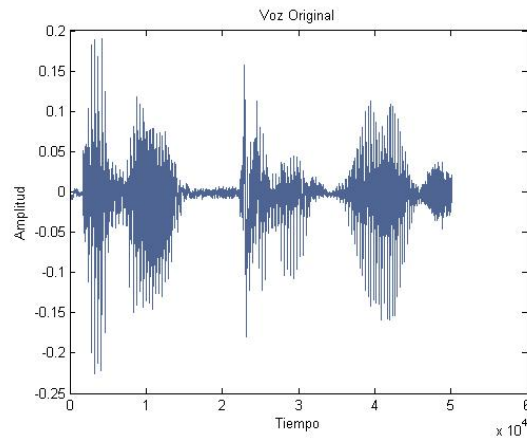
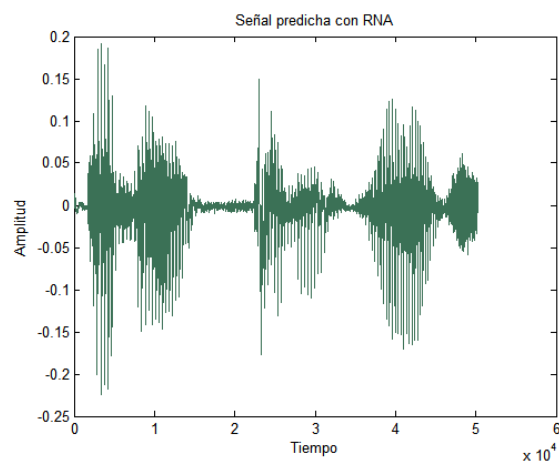
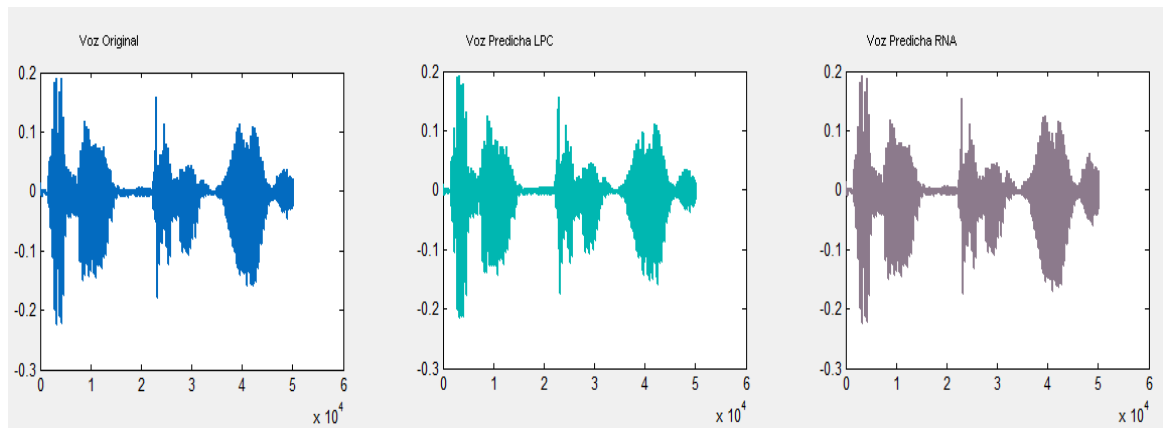


Figura 19. Señal predicha por RNA correspondiente a la frase “Aló, Buenos días”



### 6.3 RESULTADOS COMPARATIVOS DE LAS TÉCNICAS DE PREDICCIÓN

Figura 20. Comparación de resultados técnicas de predicción



Las dos técnicas de predicción arrojaron resultados satisfactorios para los objetivos propuestos, el principal de ellos: predecir la señal de voz, la percepción acústica que se logra no permite la distinción entre las frases, gráficamente (figura 20) las 3 voces indiscutiblemente se ven similares excepto en algunos lugares que son mínimos al compararlos con toda la señal en general, numéricamente los resultados son:

Figura 21. Resultados al evaluar los parámetros matemáticos generales.

Calculos Generales			
	Voz Original	Voz LPC	Voz RNA
Error cuadratico medio	0.0	8.6074e-005	2.0455e-005
Frecuencia fundamental	33.8252	29.8749	32.2077
Potencia	0.00067369	0.00059654	0.00064173
Correlacion	1	0.45698	0.99575

El error cuadrático medio se calcula al igual que los resultados obtenidos evaluando cada técnica como en la ecuación (33). Donde el error alcanzado con un exponencial de  $10^{-5}$  es excelente considerando que está muy cercano a cero.

La frecuencia fundamental corresponde a la frecuencia más baja del espectro tal que las frecuencias dominantes pueden expresarse como múltiplos de esta, como puede observarse (en la figura 21) los resultados son similares entre sí pero numéricamente hablando la frecuencia fundamental de la voz predicha mediante RNA se acerca más a la de la voz original que la de la predicción de LPC, la diferencia entre RNA y voz original es de 1,6175 hz, y entre LPC y Voz original es de 3,9503hz, entre menor sea la diferencia más cercanas están sus características entre las voces.

Potencia: la potencia de las señales de voz es una medida que también tiene proximidad entre las distintas señales para LPC el valor es de 0,00059654, con una diferencia con la original de  $7,715^{-5}$  con respecto a la original, para RNA el valor es de 0,00064173 con una diferencia de  $3,196^{-5}$ , de nuevo la diferencia con RNA es menor que con LPC.

Correlación: esta Medida se tomó con respecto a la señal original, la proximidad que tiene la voz predicha con la original, un valor cercano a 1 es aceptablemente bueno, como se puede observar la correlación que presenta la voz predicha por RNA da muy próxima a 1, mientras que la de LPC es aproximada a 0,5. La mejor medida con respecto a correlación la presenta la voz predicha mediante RNA.

Puede observarse en la tabla 11 las diferentes pruebas que se realizaron cambiando la frase (F) y el individuo (I). Los resultados permiten ver que el comportamiento de las dos técnicas es bueno para el estudio realizado, en algunos casos LPC logra superar a RNA y viceversa; RNA por su característica de

no linealidad tiene mejores aproximaciones en las características propias de la voz (Frecuencia fundamental, Potencia y coeficiente de correlación).

E2M es el error cuadrático medio entre la voz original y la voz predicha ya sea por la técnica LPC o RNA.

Tabla 11. Resultados de predicción por LPC y RNA al variar los individuos y las frases

I	F	E2M		Frecuencia Fundamental			Potencia			Coeficiente de Correlación	
		Voz LPC	Voz RNA	Voz Original	Voz LPC	Voz RNA	Voz Original	Voz LPC	Voz RNA	Voz LPC	Voz RNA
1	1	0.0010965	0.00012888	432.9686	334.7313	417.1458	0.12867	0.10357	0.12471	0.036756	0.79834
	2	0.00058003	0.00040527	449.6093	357.133	397.4263	0.14508	0.12	0.12908	0.25463	0.78345
	3	0.00078876	0.0000492	557.1095	481.2242	495.5785	0.093007	0.082176	0.083011	0.019556	0.84945
	4	0.00075055	0.00019933	1205.5261	940.5641	1031.9704	0.1658	0.13181	0.14232	0.1268	0.79184
	5	0.0006994	0.00014301	660.8474	563.8621	586.8598	0.13921	0.12194	0.12415	0.15794	0.80465
2	1	0.00066545	0.0001449	230.3424	199.8369	202.3527	0.060331	0.054068	0.053279	0.23156	0.81468
	2	0.0010575	0.00013923	175.6358	142.3136	162.1121	0.049713	0.041759	0.046146	0.30075	0.74545
	3	0.00042894	0.00064526	332.7237	298.2315	257.4493	0.051673	0.047308	0.040107	0.025024	0.82869
	4	0.00023168	0.00016933	375.6768	298.2315	368.5701	0.046662	0.047308	0.045893	0.007503	0.75348
	5	0.00011277	0.00098059	309.7788	269.9759	269.705	0.05955	0.053229	0.052047	0.29986	0.80321
3	1	0.0055565	0.00014246	880.5854	786.4588	711.4876	0.21965	0.20311	0.17836	0.57278	0.87434
	2	0.0055565	0.00014246	880.5854	786.4588	711.4876	0.21965	0.20311	0.17836	0.57278	0.87434
	3	0.0044321	0.0010321	682.0317	614.2471	568.1583	0.1064	0.097686	0.088914	0.36119	0.83417
	4	0.0085014	0.00035043	1426.2229	1249.0132	1196.8603	0.17538	0.15613	0.14754	0.25756	0.84777
	5	0.0033778	0.0015144	568.9206	512.9975	552.3188	0.13782	0.12825	0.13445	0.42976	0.95338
4	1	0.00017842	0.00022972	635.0523	494.6334	566.2638	0.17824	0.14379	0.15983	0.060164	0.65972
	2	0.00093943	0.0013514	476.0888	393.848	395.5358	0.1881	0.1641	0.15752	0.17536	0.73716
	3	0.00011634	0.00060332	527.3143	438.6598	451.4806	0.095511	0.081354	0.082072	0.0077074	0.73757
	4	0.00099607	0.00029712	1684.0565	1302.2008	1358.2546	0.22884	0.18006	0.18507	0.020947	0.66008
	5	0.0049942	0.0057732	1135.6928	935.0022	914.7617	0.2715	0.23098	0.21974	0.14562	0.70486
5	1	0.00025565	0.0019252	53.8529	41.3544	44.3975	0.014224	0.011287	0.011789	0.01576	0.82582
	2	0.00017862	0.00058433	115.3704	100.9514	97.4676	0.036649	0.033383	0.03116	0.51423	0.87017
	3	0.00021554	0.00081281	189.5602	163.2166	179.942	0.031741	0.027948	0.030232	0.065162	0.90284
	4	0.00032277	0.00092284	258.0925	213.1584	247.4627	0.035851	0.030141	0.03447	0.067388	0.85488
	5	0.00011521	0.00011272	157.8963	129.511	143.9003	0.037143	0.031374	0.034011	0.20623	0.86305

### 6.3.1 Análisis de frecuencia

Figura 22. Representación en frecuencia de la señal original

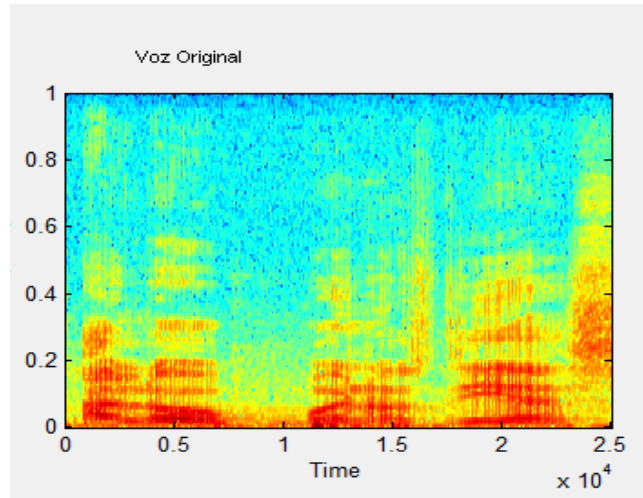
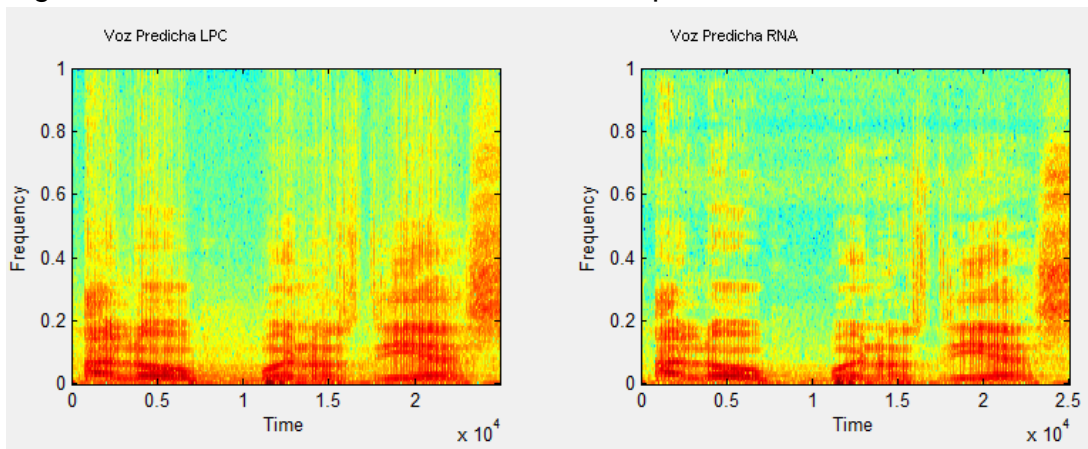


Figura 23. Análisis de frecuencia de la señal predicha

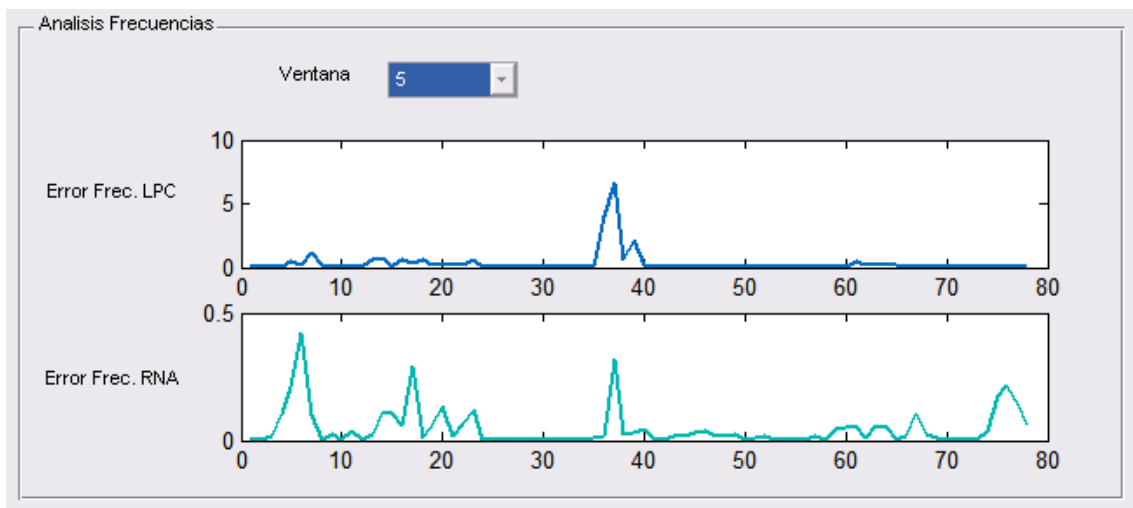


El espectrograma permite analizar la sonoridad, duración y estructura de los formantes (timbre), intensidad, pausas y ritmo, comparando los resultados gráficos se puede observar que los espectrogramas de las señales predichas con respecto a la señal original muestran que la voz predicha por RNA tiene mejores resultados que la voz predicha mediante LPC. Las frecuencias altas y bajas se encuentran aproximadamente iguales.

Para corroborar los resultados visuales y obtener un error numérico de valores de frecuencia se promedió la medida de las frecuencias mostrada en el espectrograma con ventanas de ancho variables de 1 a 10 columnas de datos.

Se demuestra que las señales tienen valores de frecuencia aproximados, en los mismos rangos de valores. Sin embargo, se presentan picos en los cuales la diferencia es considerable. En la Figura 24 se muestra el comportamiento para la frase “Hola, buenas tardes doctor”, con una ventana de ancho igual a 5 donde se pueden observar picos en la grafica de error. Además, el valor de estos picos para la señal predicha con LPC es mayor, la señal predicha con RNA tiene mayor aproximación en frecuencia con la original, como era de esperarse después de observar los cálculos de la frecuencia fundamental.

Figura 24. Análisis del error al calcular las frecuencias por ventana



### 6.3.2 Pruebas humanas

Con el fin de evaluar la validez los resultados de predicción frente al usuario final se realizaron pruebas humanas con 10 individuos. Estas pruebas consistían en reproducir 10 veces una señal variando aleatoriamente entre la señal original y la predicha. La persona decía que señal correspondía según lo que captara. Estas pruebas dan una idea general de que tanta diferencia hay entre una y otra. Sin embargo, esta medida tiene el inconveniente, después de reproducir las señales más de una vez la persona puede identificar características propias de cada una, es equivalente a lo explicado en el apartado de RNA, la interconexión de neuronas se refuerza al presentarse una y otra vez la misma señal y por lo tanto el individuo empieza a memorizar. A pesar de esta desventaja, el sistema responde satisfactoriamente puesto que en la mayoría de los casos la persona no sabía cuál de las señales estaba escuchando.

Tabla 12. Pruebas Humanas de Predicción con frases repetitivas

Ind.	LPC		RNA	
	Acierto	Error	Acierto	Error
1	2	8	9	1
2	2	8	4	6
3	4	6	3	7
4	4	6	5	5
5	5	5	4	6
6	7	3	3	7
7	4	6	3	7
8	10	0	3	7
9	4	6	10	0
10	4	6	10	0

En la tabla 12 se observa el número de aciertos que corresponde al número de veces que la respuesta de la persona fue correcta y el número de errores que

indican lo contrario, es decir, el número de veces en que la persona se equivocó. Si la persona tiene hasta 5 aciertos se puede asegurar que no es capaz de identificar las señales, si tiene hasta 7 aciertos la persona probablemente aun este adivinando, si tiene 8 o más aciertos la persona puede identificar cada una de las señal. En la tabla se observa que para el caso de LPC solamente un individuo tenia total certeza de que señal se reproducía en cada caso y otro tuvo un valor de acierto igual a 7. Para el caso de RNA 3 individuos captaron las diferencias entre la original y la predicha.

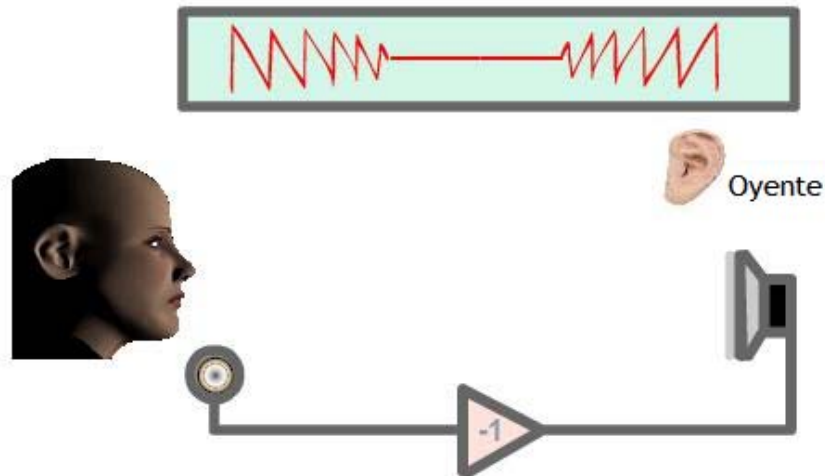
#### **6.4 CANCELACIÓN DE LA SEÑAL**

La parte final del trabajo propuesto y por ende del desarrollo del proyecto es la cancelación de la señal, en donde después de reconstruir la señal es posible realizar la cancelación.

Para atenuar la señal, la muestra predicha es invertida multiplicando los datos por -1, en seguida la señal original es enviada por un canal del parlante y la invertida por otro (cada una viaja por un amplificador diferente), los amplificadores son situados de manera estratégica, de tal forma que al enviar las dos señales al mismo tiempo la atenuación y cancelación de sonido sea posible. (Ver Figura 25)

De esta forma se pretende que la señal resultante tenga menor presión acústica (disminuir el número de decibeles), con respecto a la señal original.

Figura 25. Modelo del sistema de cancelación



La efectividad del sistema depende de la posición de los parlantes, distancia y ángulo de separación entre ellos, de igual manera depende de la ubicación del oyente en relación a la fuente de sonido, según el lugar en que se encuentre este puede captar la señal con mayor o menor presión acústica. Utilizando un sonómetro, se determinó la alineación adecuada de los parlantes y se analizó la variación del sonido dependiendo la posición del oyente, estableciendo los puntos donde se obtienen mejores resultados (captar medidas de decibeles bajas). A continuación se muestra la experiencia

#### **6.4.1 Pruebas numéricas de cancelación mediante un sonómetro**

Para llevar a cabo las pruebas se utilizó en la toma de medidas un sonómetro digital, instrumento diseñado para reaccionar al ruido como el oído humano. Las medidas tomadas por los sonómetros generalmente son comparadas a estándares conocidos para identificar si el nivel del ruido podría tener efectos molestos en humanos localizados cerca de la fuente del ruido.

Las Aplicaciones Típicas de los Sonómetros Incluyen:

- Medida Profesional de Ruido y Conformidad
- Medida Ambiental de Ruido y Conformidad
- Aplicación de Ordenanza de Ruido
- Evaluación de Ruido de Comunidad
- Estudios del Impacto Ambiental
- Análisis de Frecuencia y Calibración
- Análisis de Audiómetro y Calibración
- Los Estudios de Control de Ingeniería
- Pruebas de los Productos

El nivel de ruido es medido en decibeles y sufre variaciones según el ruido ambiental, la sensibilidad del instrumento permite percibir desde un pequeño susurro hasta ruidos demasiado molestos, subiendo los niveles en decibeles. La siguiente tabla muestra las especificaciones técnicas del sonómetro:

Tabla 13. Especificaciones técnicas del sonómetro

<b>Especificaciones Técnicas</b>	<b>SoundPro</b>
IEC/ANSI clase 1, estándar SPM	X
Decibelímetro integrador	1
Analizador de frecuencia en tiempo real	X
Análisis de frecuencia A, C, Z (lineal)	X
Valoración temporal F, S, I	X
Valor umbral 0 - 140 dB	X
Parámetro de reducción 3, 4, 5, 6 dB	X
Espectro de octava (16 Hz ... 16 kHz)	X
Espectro de tercios (12,5 Hz ... 20 kHz)	X

Pantalla y teclado iluminado	X
Conexión USB	X
Inicio y fin programable y activación de nivel	X
Registro de datos	X
Intervalo de registro de 1 s hasta 60 min	X
Mediciones disponibles: SPL, MAX, MIN, Peak, Ln, Leq, Lavg, SEL, TWA, Taktmax, DOSIS, DOSIS ampliada, Ldn, CNEL, Exposition	
Intervalo de registro: 1 , 10, 15, 30, 60 segundos 5, 10, 15, 30, 60 minutos	

Figura 26. Sonómetro Digital Sound Pro



Empleando el dispositivo descrito (Ver figura 26) se procede a escoger la mejor posición de los parlantes y del oyente, describir el ambiente y las medidas a calcular.

#### 6.4.1.1 Escoger la mejor posición del parlante

Se realizaron experimentos de medida de potencia, el procedimiento consistió en tomar un punto fijo para el receptor a una distancia de 50 cm frente a los parlantes, en este punto se ubicó el sonómetro para medir los decibeles percibidos en esas condiciones mientras se emitía un sonido modificando la posición del segundo parlante. (Como puede observarse en las imágenes de la figura 27), las mediciones dieron lugar para escoger una buena posición de la fuente al locutor.

Figura 27. Variación de los parlantes



Parlante 1: Estático, equivale al emisor de la señal

Parlante 2: Variable, equivale a la fuente de cancelación.

La unidad de sonido empleada para tomar las muestras es decibeles (dB) la cual expresa la relación entre las presiones de un sonido cualquiera y un sonido de referencia en escala logarítmica. Equivale a 20 veces el logaritmo de base 10 del cociente de las dos presiones.

Numéricamente los resultados se visualizan en la tabla 14:

Tabla 14. Potencia según la disposición de los parlantes

Separación entre parlantes (cm)	Altura (cm)	Descripción de la posición	dB	dB Disipados
3	0	Totalmente de frente	47,9	14,1
10	0	Totalmente de frente	49,2	12,8
20	0	Totalmente de frente	50,2	11,8
3	10	El 2 parlante está en forma horizontal, sobre una base	49	13
5	0	El 2 parlante está sobre el plano en forma horizontal	53,7	8,3
15	0	A 36 grados de separación	50,3	11,7
6	0	38 grados de separación, el parlante se encuentra de espalda al oyente	54	8
6	0	38 grados de separación, parlante de frente al oyente	52,4	9,6

Sin una fuente disipadora de potencia el sonómetro indicó que un oyente en la misma posición percibía un valor de 62 dB. (Valor equivalente a una conversación normal), con un ruido ambiental de 46 dB. Los dB disipados se refieren a la potencia normal de la voz sin cancelar menos la potencia de la voz cancelada, se observa cuantos decibeles logra disminuir el sistema.

Como puede observarse en la tabla, el sistema de cancelación definitivamente disminuye los decibeles percibidos, pero la mejor posición es a 3 cm de separación totalmente en frente del emisor, se obtiene una potencia de 47,9 dB, comparable con la intensidad normal de 62 dB, disminuye 14, 1 dB la potencia de la señal 22,9% del valor inicial, resultado favorable para el propósito de cancelación.

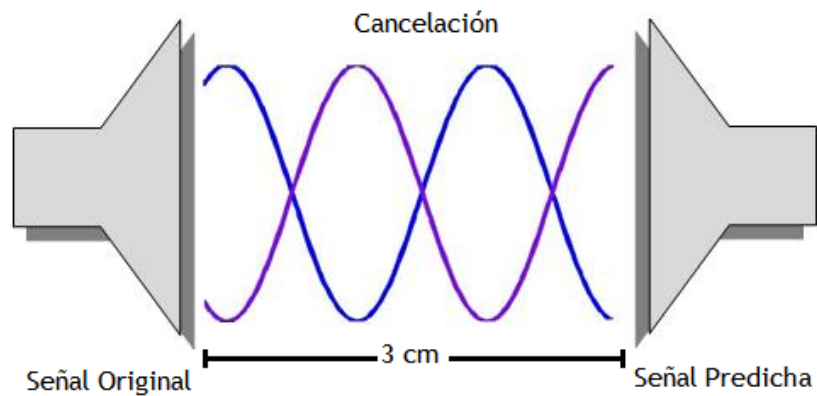
Se puede observar la imagen del montaje final en la figura 28 y el diseño de la figura 29.

Figura 28. Configuración final de los parlantes (3 cm de separación)



Los parlantes se encuentran a 3 cm de distancia, el sonómetro es ubicado en diferentes puntos del plano para tomar las potencias

Figura 29. Diseño final del modelo de cancelación.

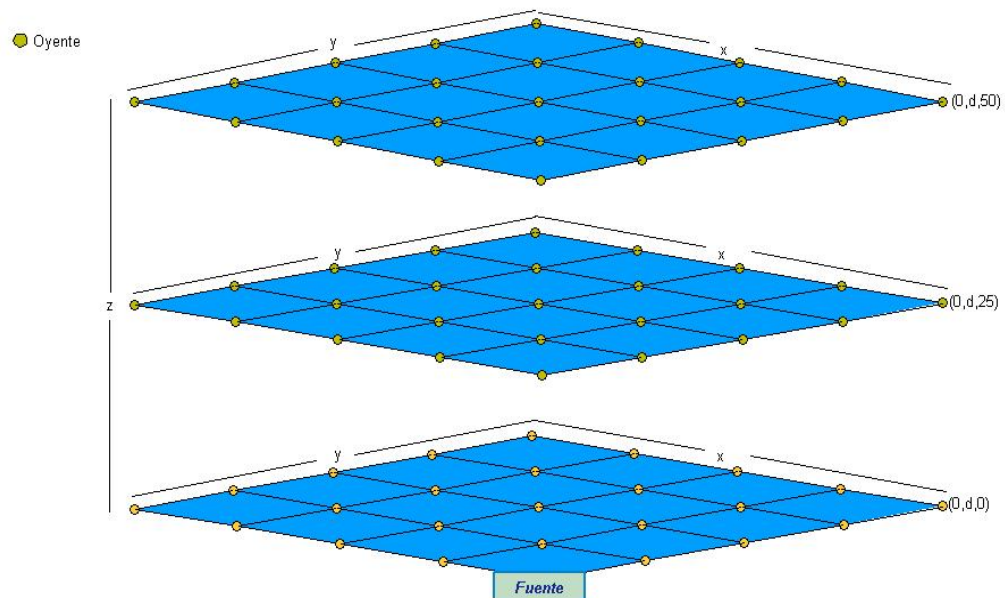


#### 6.4.1.2 Posición del oyente externo

Para variar la posición del espectador se realizó el experimento en dos entornos diferentes. Las primeras valoraciones se realizaron en un plano de  $132 \text{ cm}^2$  variando distancias en  $x$  y cada  $33 \text{ cm}$ . en tres diferentes alturas ( $0, 25$  y  $50 \text{ cm}$ .) para un total de  $25$  medidas por plano y  $75$  datos en total. El segundo entorno se hizo con un plano de  $2 \text{ m}^2$ , en este caso la variación de las distancias en  $x$  y  $y$  es de  $50 \text{ cm}$ . de  $0$  a  $200 \text{ cm}$ . Igualmente, en el caso de la altura ( $z$ ) se tomaron valores de  $0$  hasta  $50 \text{ cm}$ . Aumentando cada  $25 \text{ cm}$ .

En la figura 30 se muestra el modelo de los planos y los puntos de ubicación del oyente externo, en donde  $d$  puede tomar el valor de  $132 \text{ cm}$  o  $200 \text{ cm}$  según el plano empleado, los puntos amarillos corresponden a las posiciones en que varía el ambiente.

Figura 30. Modelo de la superficie para la toma de presión acústica



Las figuras 31 y 32 ilustran el experimento para los dos entornos en posiciones particulares:

Figura 31. Experimento en el plano de 132cm<sup>2</sup>.

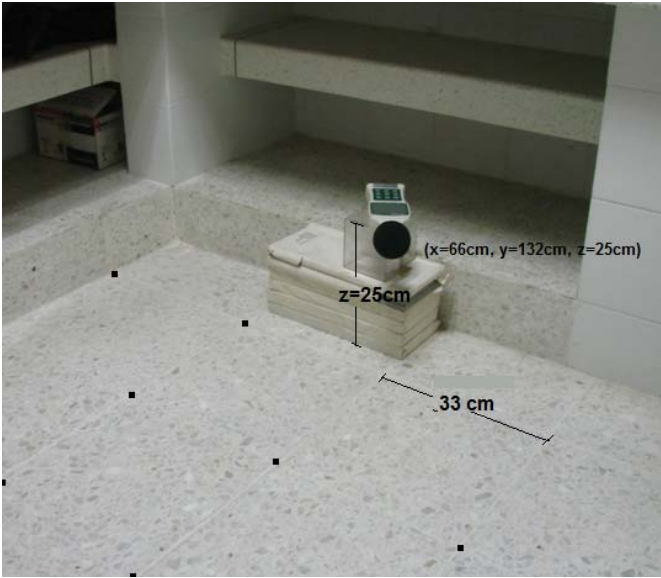
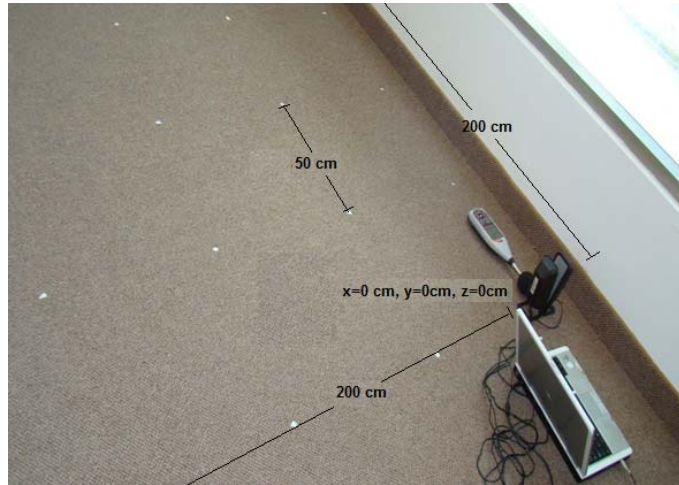


Figura 32. Experimento en el plano de 200 cm<sup>2</sup>



### 6.4.1.3 Ambiente para la toma de muestras

Las potencias capturadas para los resultados se tomaron en dos tipos escenarios diferentes (figura 33):

Figura 33. Escenarios de las pruebas



- Sala de grabación CENTIC Universidad Industrial de Santander, con un ruido ambiental variable entre 45 y 47 dB.

- Laboratorio de acústica Fiscalía general de la Nación de Bucaramanga, con un ruido ambiental variable entre 30 y 33 dB.

#### **6.4.1.4 Medidas a tomar para el experimento**

Para este primer experimento, que consiste en enviar la señal original por un canal y la señal desfasada 180 grados por el otro canal y calcular en los diferentes puntos del plano dos valores de potencia:

- La potencia que es emitida normalmente sin la fuente disipadora, es decir, la potencia de emisión de la señal normal.
- La potencia al poner en funcionamiento el sistema de cancelación con la señal original e invertida, es decir, la potencia emitida cuando existe una fuente externa que emite la señal invertida.

#### **6.4.2 Resultados del experimento**

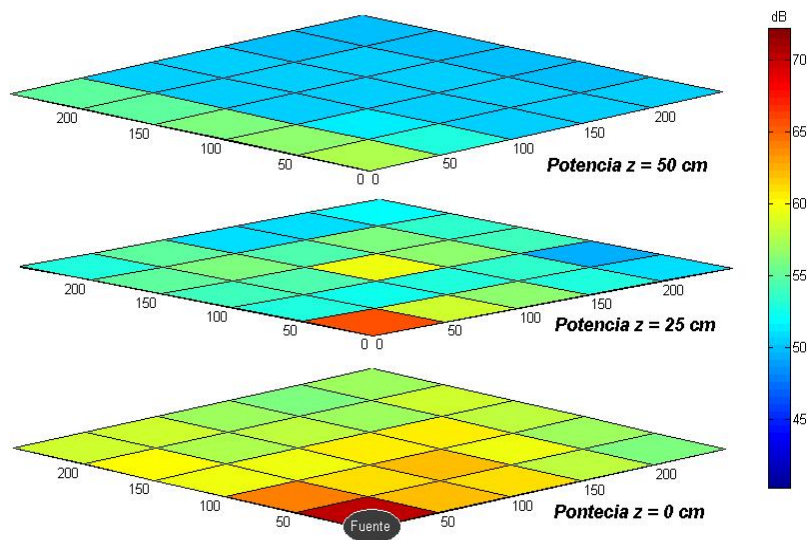
##### **6.4.2.1 Escenario 1: Ruido ambiental a 46 dB**

Como se comentó en el apartado anterior en este entorno se tomaron medidas en un plano de 2 m<sup>2</sup>, se visualiza el comportamiento resultante del sistema de cancelación en la tabla 16 en donde se resumen las mediciones de potencia para el plano en z=0, la posición hace referencia a el oyente externo, es decir, la posición del sonómetro para calcular los valores percibidos en las diferentes posiciones. Tomando en consideración la posición de la fuente la cual se encuentra a la misma altura.

Tabla 15. Mediciones de potencia para el plano de 0 a 200 cm modificando la posición cada 50 cm, ruido ambiental promedio de 46 dB

Posición en x	Posición en y	Potencia sonora voz normal (dB)	Potencia sonora voz cancelada (dB)	Potencia atenuada (dB)
0	0	70,3	60	10,3
0	50	62	52,7	9,3
0	100	61	49,3	11,7
0	150	58,1	48	10,1
0	200	56	45,8	10,2
50	0	64,1	56,7	7,4
50	50	61,1	49,9	11,2
50	100	61,9	46,8	15,1
50	150	59,6	47,6	12
50	200	57,1	46,2	10,9
100	0	59,7	55,3	4,4
100	50	59,4	48,8	10,6
100	100	60,6	50	10,6
100	150	60,7	50,4	10,3
100	200	58,2	48,6	9,6
150	0	60,2	48,9	11,3
150	50	57,4	47	10,4
150	100	58,2	48,7	9,5
150	150	57,2	46,9	10,3
150	200	58,7	47,3	11,4
200	0	58,7	49,1	9,6
200	50	58,3	47,9	10,4
200	100	56,9	47,3	9,6
200	150	56,1	46,6	9,5
200	200	57,2	47,3	9,9

Figura 34. Potencia de la señal sin cancelar a volumen habitual. Escenario CENTIC

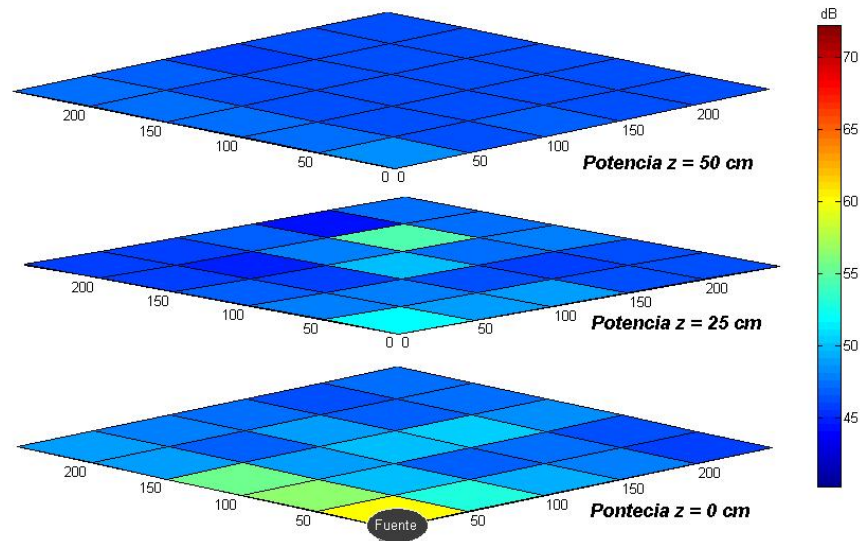


Para tener una idea de la potencia de una señal normal, cuando un individuo habla habitualmente se tomaron los valores en los mismos planos de la señal emitida a un volumen promedio por un orador. En la figura 34 se observan los valores de potencia en los tres planos considerados.

Siendo los valores de menor potencia la gama de colores fríos, el menor valor es igual a 40 dB, el cual corresponde al color azul oscuro. Los valores de mayor potencia se relacionan con los colores cálidos, el máximo valor percibido es el color rojo intenso el cual corresponde a 70.3 dB de potencia. Este valor se obtiene justo en la posición en la cual se sitúa la fuente, es decir, corresponde a la potencia de la voz del emisor. Al aumentar la altura del oyente la potencia disminuye como se observa gráficamente, de donde se concluye que la potencia es inversamente proporcional a la distancia.

Una vez puesto en funcionamiento el sistema de cancelación la potencia en todos los planos disminuye considerablemente como es reflejado en la figura 35.

Figura 35. Potencia de la señal cancelada. Escenario 1



Al comparar esta imagen con la anterior es evidente que el sistema logra disipar hasta los niveles más bajos de potencia del sonido en el área circundante a un orador cumpliendo los objetivos propuestos en el proyecto.

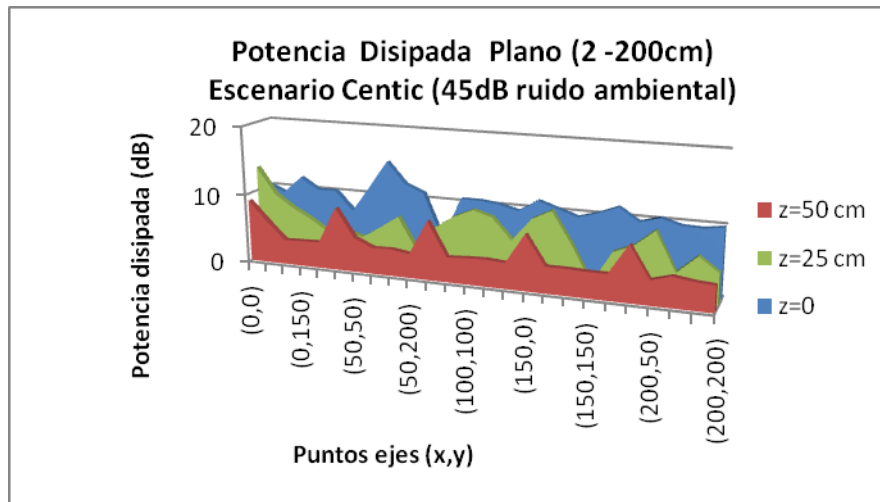
En promedio se logra disminuir en 10dB el ruido causado por la contaminación del discurso activo, lo cual es observable en la última columna de la tabla 15, potencia atenuada (dB), la cual hace referencia a la resta entre la potencia de la señal normalmente y la potencia después de cancelar.

A partir de ciertos puntos en que el oyente se encuentra más alejado de la fuente el sonido percibido tiene los mismos niveles del ruido ambiental.

Para observar los valores de las potencias percibidas en iguales condiciones y en el mismo plano pero a alturas diferentes ( $z = 25 \text{ cm}$  y  $z = 50 \text{ cm}$ ). Ver ANEXO B Escenario 1.

La figura 36 permite observar la potencia disipada o atenuada en cada plano en que se tomaron las medidas, en la grafica el plano de color azul corresponde a la casilla de potencia atenuada de la tabla 15. Se puede observar que para estos planos la potencia disipada, el valor es más pequeño y poco constante debido a que en ciertos valores el dispositivo percibe el sonido como ruido ambiental por las condiciones del lugar estos valores oscilan un poco más.

Figura 36. Potencia Disipada Plano (0 -200cm) Escenario CENTIC ( 45dB ruido ambiental)



#### 6.4.2.2 Escenario 2: Ruido ambiental a 33 dB.

Para tener otra perspectiva de medidas y resultado del sistema de cancelación se realizaron pruebas equivalentes a las del escenario 1 pero en un ambiente con menor intensidad de ruido ambiental y en este caso las medidas se tomaron en un área de 132 cm<sup>2</sup>. Estas pruebas se realizaron en el laboratorio de acústica del CTI\* de la fiscalía. En la tabla 16 se exponen los valores derivados de las pruebas.

Tabla 16. Mediciones de potencia para el plano de 0 a 132 cm con variación de 33 cm en cada punto, promedio de ruido ambiental 33 dB.

Posición en X	Posición en y	Potencia sonora voz normal (dB)	Potencia sonora voz cancelada (dB)	Potencia atenuada (dB)
0	0	72,5	57,2	15,3
0	33	65,8	51,7	14,1

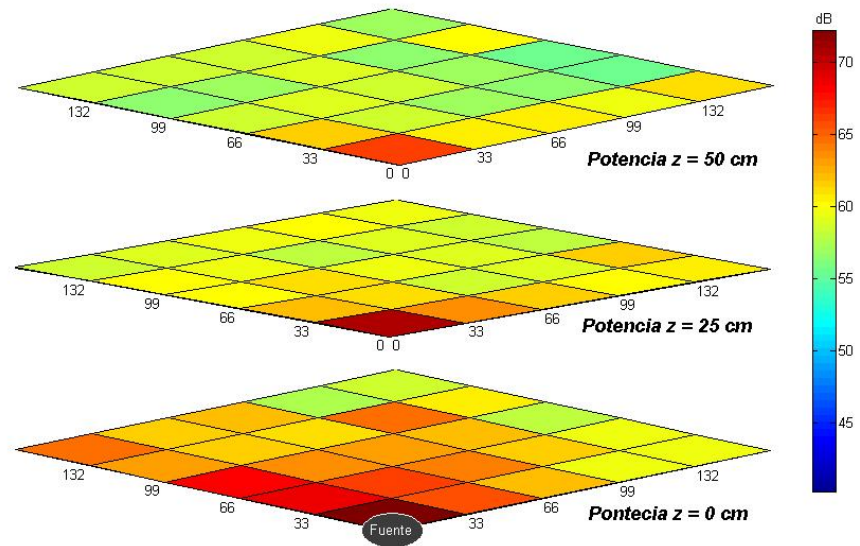
\* CTI: Cuerpo Técnico de Bucaramanga

0	66	62,3	49,9	12,4
0	99	59,6	47,1	12,5
0	132	59,5	46,6	12,9
33	0	68,9	55,5	13,4
33	33	66	51	15
33	66	64	50,2	13,8
33	99	61,4	47,2	14,2
33	132	59,7	46,4	13,3
66	0	68,2	52,8	15,4
66	33	63,6	49,1	14,5
66	66	63,1	49,8	13,3
66	99	61,9	47,2	14,7
66	132	58	48,3	9,7
99	0	63	49,1	13,9
99	33	61,6	49,6	12
99	66	61,1	46,4	14,7
99	99	64,4	48,8	15,6
99	132	60,5	47,9	12,6
132	0	64,6	51,4	13,2
132	33	61,5	49,5	12
132	66	62,2	50,4	11,8
132	99	57,7	48,8	8,9
132	132	58,4	45,5	12,9

La potencia de la señal de la voz normal en este entorno en cada punto de los planos es la mostrada en la figura 37.

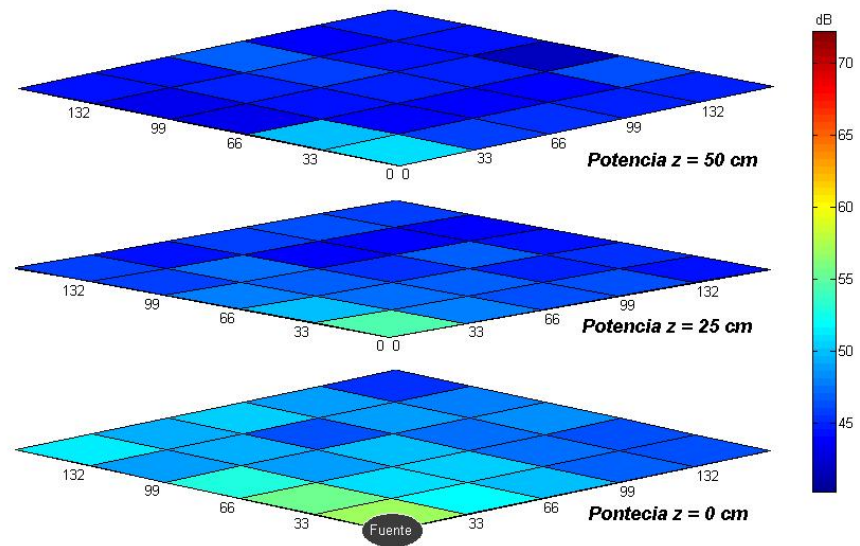
Debido a que el ruido ambiental es menor en este entorno se estiman mejor los planos de percepción de la potencia sonora en cada punto en el cual se ubico el oyente, como puede observarse en los puntos más próximos de la fuente sonora se presentan los mayores valores de potencia. El valor de rojo intenso hace referencia al valor de mayor potencia (73 dB), es decir, donde se encuentra la fuente sonora. Como era de esperarse en los planos de 25 y 50 cm la potencia es menor.

Figura 37. Potencia de la señal a volumen habitual. Escenario 2



Para comparar la potencia de la señal normal con la potencia resultante del sistema de cancelación se muestra la figura 38:

Figura 38. Potencia de la señal cancelada. Escenario 2

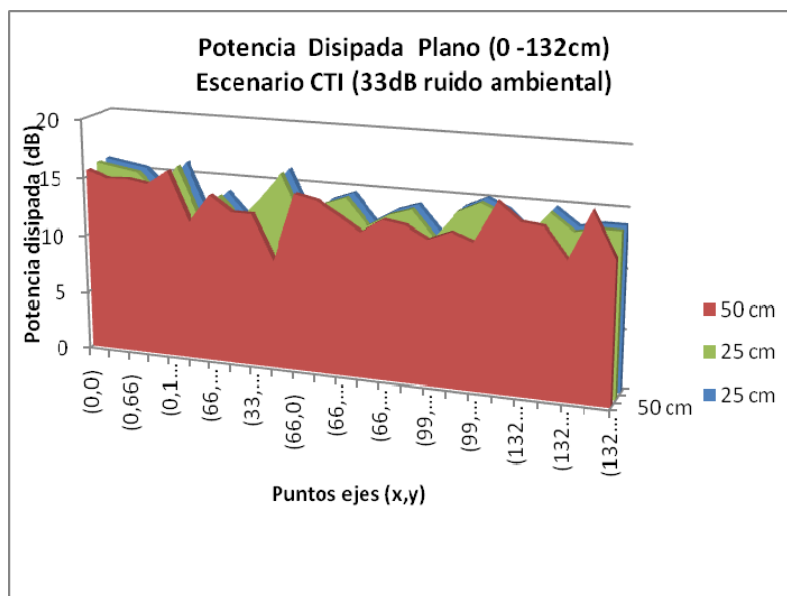


Es evidente los buenos resultados que arroja el sistema de cancelación, puesto que los valores de potencia disminuyen notablemente pasando de tonos rojos (máximo valor) a tonos azules (mínimo valor) en donde la potencia más pequeña es igual a 43 dB.

En promedio el ruido de la fuente emisora se disminuye en 13,6 dB cuando tiene la fuente disipadora en frente, debido a que el ruido ambiental es más bajo se logra percibir mejor la disminución del ruido causado por el discurso activo.

La disminución de potencia o potencia atenuada para este plano, se aprecia en la figura 39. En donde se muestra esta medida en cada plano. Se puede observar que la potencia disipada es relativamente invariable para los 3 planos, debido a que este escenario tiene el ruido ambiental más bajo se puede tener una aproximación constante del nivel disipado, aproximadamente 13 dB.

Figura 39. Potencia Disipada Plano (0 - 132 cm) Escenario CTI (33 dB ruido ambiental)



Para un análisis detallado de las potencias disipadas en los planos  $z=25$  y  $z=50$ . Ver ANEXO B Escenario 2.

Con estas pruebas se demuestra que la cancelación de sonido es posible dependiendo la posición de los elementos involucrados (parlantes y oyentes). Se percibe el grado de atenuación en el área circundante al emisor, disminuyendo notablemente la potencia comparado con los valores de la señal original sin el disipador.

La potencia más alta en cualquier caso es cuando el oyente está sobre los parlantes (posiciones cero, con alturas 0 cm, 25 cm y 50 cm) equivale a la potencia que escucharía el mismo orador.

A medida que el oyente externo se aleja del origen como era de esperarse, el valor de la potencia disminuye en relación a la distancia, por tanto, se puede observar que el plano de menor intensidad en general corresponde a  $z= 50$  cm., la persona está muy lejos de la fuente por lo cual la potencia es bastante pequeña. Después de desplazarse en  $x$  y  $y$  un valor mayor o igual a 50 cm. la variación en dB no es considerable.

#### **6.4.2.3 Resultados finales de las pruebas numéricas**

Las condiciones ideales para probar la atenuación de la señal en el sistema de cancelación sería un lugar con una potencia ambiental igual a 0 dB. Las pruebas se hicieron en ambientes controlados donde no había ruido externo y la potencia era mayor a cero. Sin embargo, el ruido ambiental en los escenarios utilizados fue constante, por lo cual se puede tomar como un parámetro base y restar este valor para mirar la potencia real emitida en cada punto de los diferentes puntos del

plano. En teoría para poder afirmar que la cancelación es efectiva el valor de la potencia de la señal debe ser equivalente a un susurro es decir menor o igual a 20 dB.

Los cálculos de las potencias tomadas tienen valores entre 24.2 y 12.5, esto se evidencian en la siguiente tabla:

Tabla 17. Mediciones de potencia para el plano de 0 restando el ruido ambiental

Posición en X	Posición en y	Pt SN	Pt SC
0	0	39.5	24.2
0	33	32.8	18.7
0	66	29.3	16.9
0	99	26.6	14.1
0	132	26.5	13.6
33	0	35.9	22.5
33	33	33	18
33	66	31	17.2
33	99	28.4	14.2
33	132	26.7	13.4
66	0	35.2	19.8
66	33	30.6	16.1
66	66	30.1	16.8
66	99	28.9	14.2
66	132	25	15.3
99	0	30	16.1
99	33	28.6	16.6
99	66	28.1	13.4
99	99	31.4	15.8
99	132	27.5	14.9
132	0	31.6	18.4
132	33	28.5	16.5
132	66	29.2	17.4
132	99	24.7	15.8
132	132	25.4	12.5

Pt SN y Pt SC corresponden a las potencias de la señal normal y la señal cancelada respectivamente después de restar el ruido ambiental, variando la posición en el plano en los puntos  $[x, y]$ .

La potencia de la señal cancelada en todas las posiciones mostradas en la tabla anterior tiene valores menores que 20 dB concluyendo que es posible realizar la cancelación del discurso activo para el área circundante emitidas por un individuo. Excepto dos posiciones sobre el parlante ( $[0,0]$  y  $[30, 0]$ ). Estas posiciones serian una ubicación imposibles para un oyente externo puesto que es la misma posición de la fuente de sonido. Se obtuvieron resultados equivalentes para el caso de los planos en  $z = 25$  y  $z = 50$ . (Ver ANEXO C)

#### **6.4.3 Pruebas cualitativas de Cancelación con usuario final.**

Como segunda medida, se tomaron pruebas humanas, teniendo en cuenta que es importante la percepción del oyente frente al sistema de cancelación debido a que éste es el usuario final y puede decidir si el sistema reduce el valor del ruido y si es un aporte para el problema propuesto. Para realizar estas pruebas se procede de la siguiente manera:

- a. Se realiza el montaje (Figura 40)

Figura 40. Parlantes totalmente de frente con una separación de 3cm



- b. Se ubica la persona a una distancia de 50 cm del montaje, que es la posición más próxima en que puede estar en una reunión
- c. Cada uno de los 20 individuos escucha 10 frases del sistema de cancelación.
- d. Por cada frase se pregunta qué percepción tenían, qué se logra escuchar, palabras entendidas o si definitivamente sólo se escuchó un murmullo o nada.

Los resultados de estas pruebas son los siguientes:

De cada 10 frases, que escuchaba cada persona, sólo 2 de los 20 individuos entendieron algunas palabras correspondientes a 2 frases. El resto de las personas encuestadas escuchaban un murmullo cuando se les mostraba el montaje de cancelación, la mayoría lograba identificar si el individuo que estaba hablando correspondía a un hombre o una mujer por el timbre de voz, las pruebas se realizaron en un salón abierto sin ruido, al modificar las condiciones, agregando ruido proveniente del exterior del salón, aumentaba la dificultad para entender el audio.

Las personas encuestadas definitivamente notaron la diferencia entre la señal emitida normalmente y cuando se cancelaba, la conclusión a la que llegaban era que se le había bajado el volumen al grado que sólo se escuchaba un murmullo y no se podía entender la frase. Se preguntó si el sonido resultante era molesto y para todas las personas encuestadas era indiferente, puesto que era casi imperceptible. (Ver ANEXO D)

## 6.5 PREDICCIÓN Y CANCELACIÓN DE LA SEÑAL

Una vez realizados los experimentos de cancelación y obtenidos los resultados favorables en cuanto a la disipación del sonido con una fuente externa en donde se envía la onda inversa a la señal original, se procede a integrar todo el sistema, la etapa de predicción y la etapa de cancelación. Para probar la efectividad del resultante es preciso medir la potencia de cancelación cuando la onda inversa equivale a la onda predicha por los métodos estudiados con anterioridad (LPC y RNA).

Las pruebas consisten básicamente en las mismas de cancelación óptima (cancelación empujando la señal original desfasada 180 grados), pero en este caso la onda invertida es la señal predicha por LPC o RNA.

En este caso, solamente se tomó el escenario de ruido externo de 33dB aproximado y se realizaron las pruebas igualmente en planos de variaciones de 33cm de 0 a 132. Con alturas de  $z=0$ ,  $z=25$  y  $z=50$ .

La tabla 18. Resume las potencias de la cancelación empujando las señales predichas por las dos técnicas, es decir, emitir por un canal la señal normal y por el otro la señal predicha invertida. Como parámetro se utiliza la potencia de la cancelación óptima para tener un estimativo de la cancelación y de la disipación que se logra, la tabla equivale a el plano  $z=0$ . Para ver los resultados en los otros dos planos dirigirse al ANEXO E

Tabla 18. Mediciones de potencia de cancelación con técnicas

Posición en X	Posición en y	Voz sin Cancelar	Cancelación óptima	Cancelación Pred. LPC	Cancelación Pred. RNA
0	0	72,5	57,2	60,6	58,6
0	33	65,8	51,7	52	51,7
0	66	62,3	49,9	49,9	49,1
0	99	59,6	47,1	47,6	47,1
0	132	59,5	46,6	46,8	46,3
33	0	68,9	55,5	55,6	56,3
33	33	66	51	51,6	51,2
33	66	64	50,2	50,2	50
33	99	61,4	47,2	47,3	47,6
33	132	59,7	46,4	46,6	46,2
66	0	68,2	52,8	53,4	52
66	33	63,6	49,1	49,3	49,9
66	66	63,1	49,8	48,8	49,4
66	99	61,9	47,2	47,9	47,3
66	132	58	48,3	47,8	46,1
99	0	63	49,1	49,2	48,8
99	33	61,6	49,6	49	49,4
99	66	61,1	46,4	46,1	46,1
99	99	64,4	48,8	48,2	48,3
99	132	60,5	47,9	47,5	46,3
132	0	64,6	51,4	51,7	50,8
132	33	61,5	49,5	48,3	48,7
132	66	62,2	50,4	51,1	50,5
132	99	57,7	48,8	48,8	48,1
132	132	58,4	45,5	45,8	46,2

La representación grafica de los valores obtenidos de potencia en el plano para la cancelación con señal predicha por LPC y RNA se muestran en las figuras 41 y 42 respectivamente. Para comparar con la potencia de la voz sin cancelar y de la cancelación óptima ver las figuras 37 y 38.

Figura 41. Potencia de la señal cancelada predicha con LPC. Escenario 2

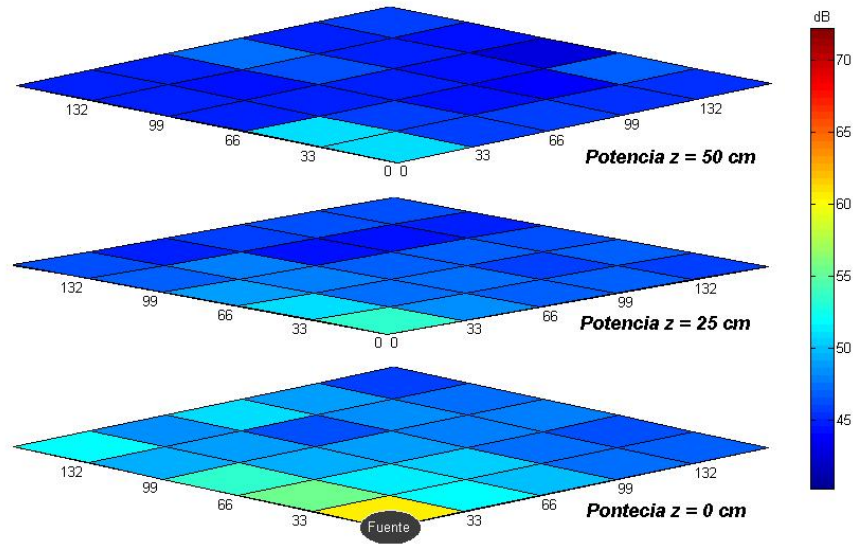
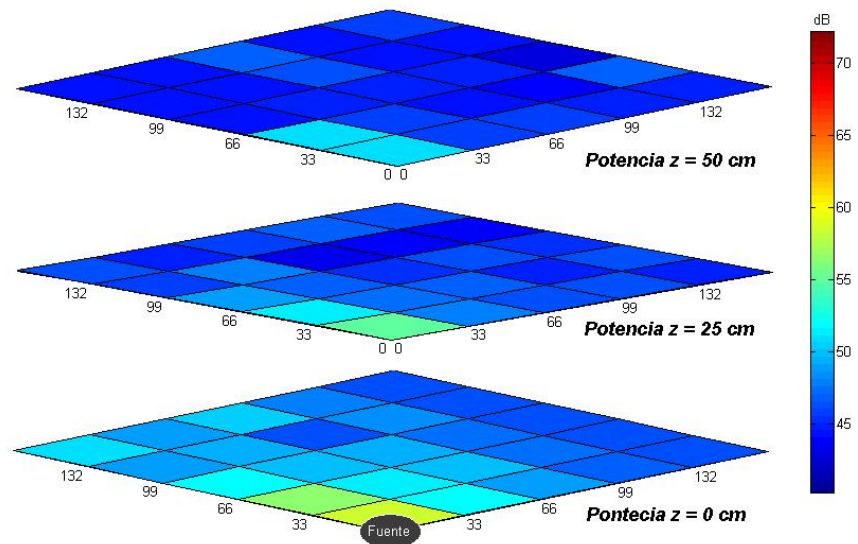


Figura 42. Potencia de la señal cancelada predicha con RNA. Escenario 2



Las gráficas de potencia para la cancelación de las señales predichas por las dos técnicas son semejantes entre si y de igual forma tienen valores en el mismo

rango a los obtenidos en la cancelación simple. Por lo cual las técnicas de predicción y cancelación se pueden integrar correctamente teniendo como resultado el “sistema de cancelación de señales de audio emitidas por un individuo en sus áreas circundantes”. Donde se hace posible predecir la señal emitida por un individuo y atenuarla a su alrededor para evitar que la señal sea percibida por un oyente externo.

Para examinar en qué porcentaje es atenuada la señal, la tabla 19 permite analizar la potencia atenuada en dB de las tres formas de cancelación (simple, LPC y RNA), en el plano  $z=0$ . La atenuación se calcula restando la potencia de la voz normal menos la potencia de la voz cancelada simple o por alguna técnica.

Tabla 19. Potencia atenuada en dB de las tres formas de cancelación

Puntos (x,y)	VN – CS	VN – LPC	VN-RNA
(0,0)	15,3	11,9	13,9
(0,33)	14,1	13,8	14,1
(0,66)	12,4	12,4	13,2
(0,99)	12,5	12	12,5
(0,132)	12,9	12,7	13,2
(33,0)	13,4	13,3	12,6
(66,33)	15	14,4	14,8
(33,66)	13,8	13,8	14
(33,99)	14,2	14,1	13,8
(33,132)	13,3	13,1	13,5
(66,0)	15,4	14,8	16,2
(66,33)	14,5	14,3	13,7
(66,66)	13,3	14,3	13,7
(66,99)	14,7	14	14,6
(66,132)	9,7	10,2	11,9
(99,0)	13,9	13,8	14,2
(99,33)	12	12,6	12,2

(99,66)	14,7	15	15
(99,99)	15,6	16,2	16,1
(99,132)	12,6	13	14,2
(132,0)	13,2	12,9	13,8
(132,33)	12	13,2	12,8
(132,66)	11,8	11,1	11,7
(132,99)	8,9	8,9	9,6
(132,132)	12,9	12,6	12,2
<b>Promedio de atenuación de Potencias</b>	<b>13,284</b>	<b>13,136</b>	<b>13,5</b>

VN – CS: Potencia voz normal – Potencia cancelación Simple

VN – LPC: Potencia voz normal - Potencia cancelación de la señal predicha con LPC

VN – RNA: Potencia voz normal – Potencia cancelación de la señal predicha con RNA

Como puede observarse para el plano  $z=0$ , el promedio de atenuación de las potencias es del orden de los 13 dB, sin embargo la cancelación de sonido tiene mayor ganancia con la técnica de predicción mediante RNA. Se logra ver una diferencia de 13, 5 dB entre la potencia de la voz normal y la potencia de la voz cancelada por RNA.

Para un mayor análisis de la atenuación de la potencia se presentan gráficamente estos resultados, la figura 43. Permite observar el plano  $z=0$ , la figura 44 el plano  $z = 25$  cm y la figura 45 el plano  $z=50$ cm.

Figura 43. Atenuación de potencia, z=0

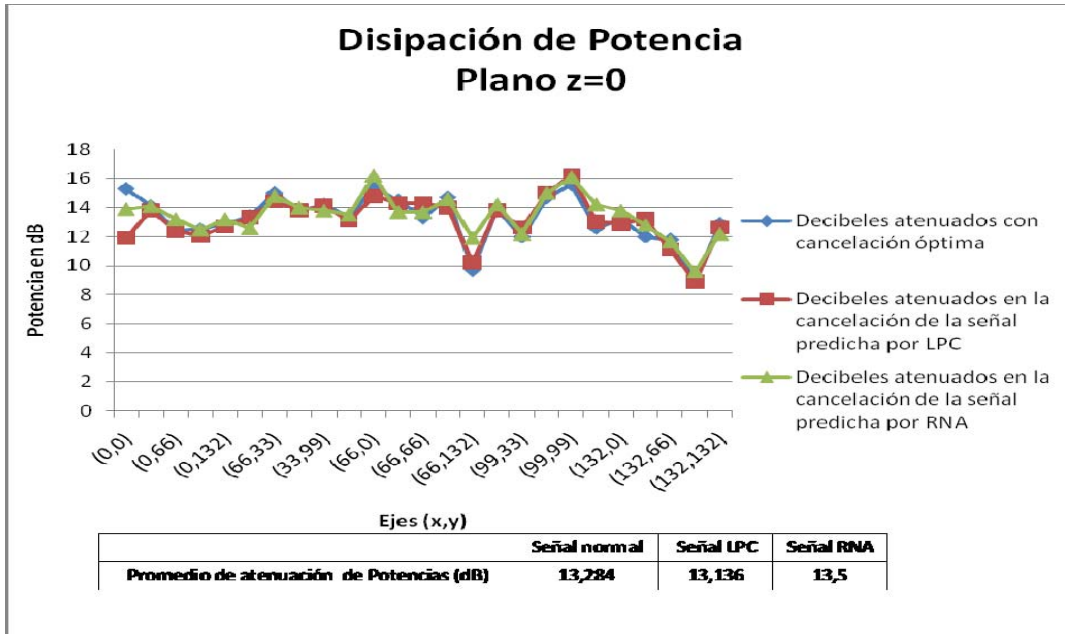


Figura 44. Atenuación de potencia, z=25

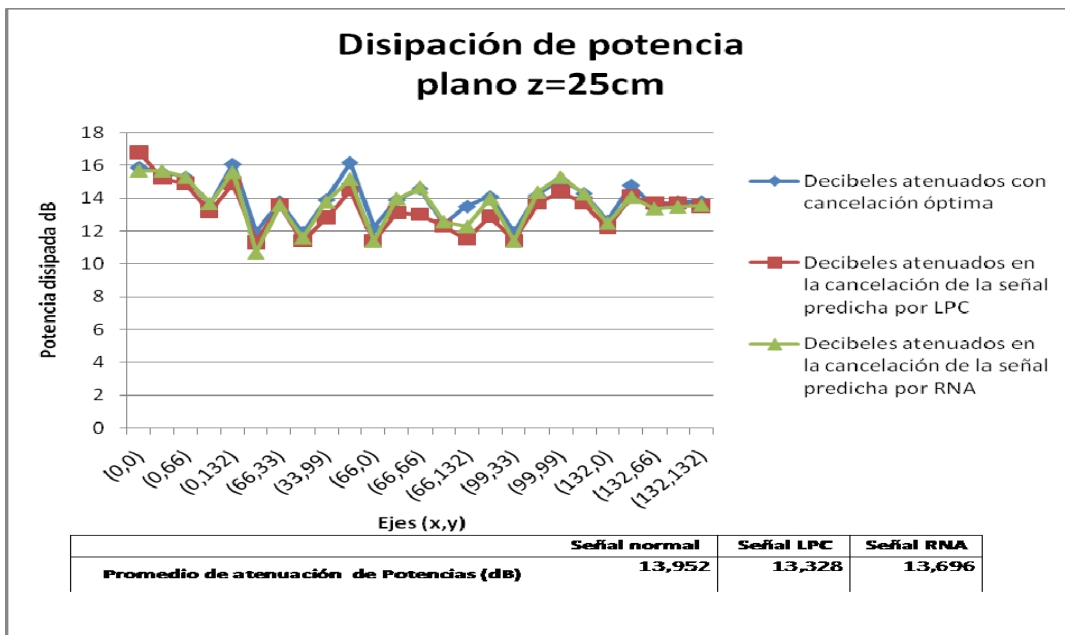
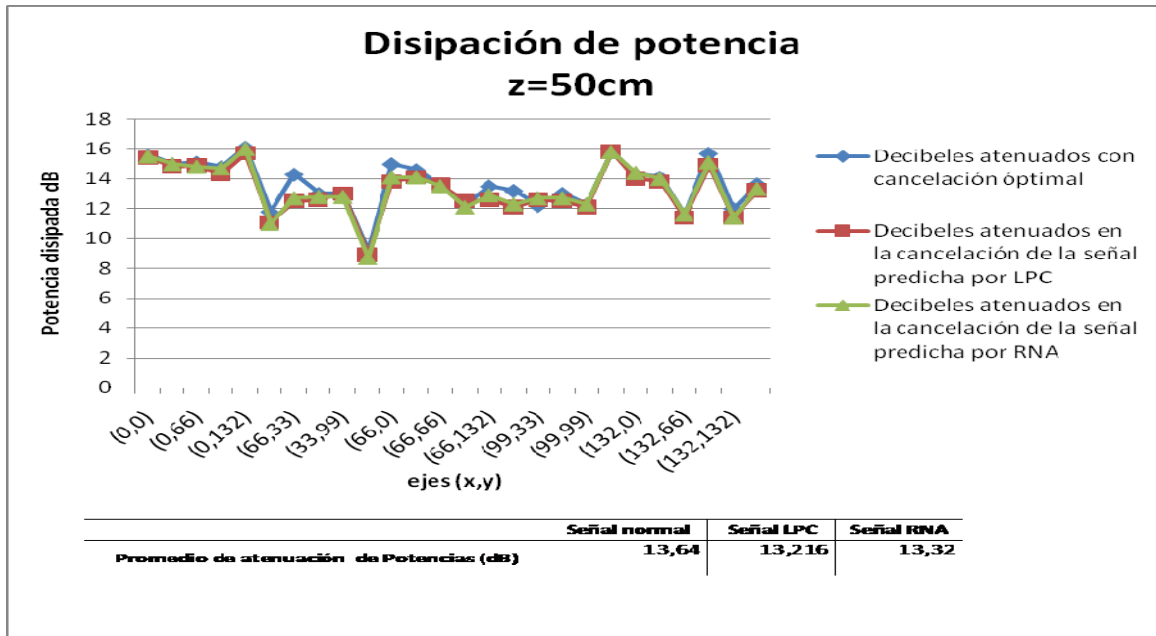


Figura 45. Atenuación de potencia, z=50



En los tres planos en promedio la potencia atenuada equivale a 13dB. Comparando las técnicas LPC y RNA aunque no es una diferencia significativa RNA logra disipar un poco más la señal de voz que LPC.

Son indudables los resultados positivos que demuestra el proceso de predicción y cancelación al considerar las tablas y graficas mostradas anteriormente. Finalmente se consideran los datos obtenidos después sustraer la potencia del ruido ambiental, haciendo un procedimiento equivalente al de la cancelación simple tenemos los siguientes resultados.

Tabla 20. Potencias sin ruido ambiental

Posición en X	Posición en y	Pt SN	Pt SC	Pt. SLPC	Pt. SRNA
0	0	39.5	24.2	27.6	25.6
0	33	32.8	18.7	19	18.7

0	66	29.3	16.9	16.9	16.1
0	99	26.6	14.1	14.6	14.1
0	132	26.5	13.6	13.8	13.3
33	0	35.9	22.5	22.6	23.3
33	33	33	18	18.6	18.2
33	66	31	17.2	17.2	17
33	99	28.4	14.2	14.3	14.6
33	132	26.7	13.4	13.6	13.2
66	0	35.2	19.8	20.4	19
66	33	30.6	16.1	16.3	16.9
66	66	30.1	16.8	15.8	16.4
66	99	28.9	14.2	14.9	14.3
66	132	25	15.3	14.8	13.1
99	0	30	16.1	16.2	15.8
99	33	28.6	16.6	16	16.4
99	66	28.1	13.4	13.1	13.1
99	99	31.4	15.8	15.2	15.3
99	132	27.5	14.9	14.5	13.3
132	0	31.6	18.4	18.7	17.8
132	33	28.5	16.5	15.3	15.7
132	66	29.2	17.4	18.1	17.5
132	99	24.7	15.8	15.8	15.1
132	132	25.4	12.5	12.8	13.2

Pt. SLPC, es la potencia de cancelación de la señal predicha con la técnica de LPC. Pt. SRNA, es la potencia de cancelación de la señal predicha con la técnica de RNA

Los valores de potencia de la cancelación de las señales predichas por las dos técnicas estudiadas están en un intervalo que en promedio no supera el equivalente a los decibeles de un susurro (20 dB), lo cual indica que si hay cancelación. Nuevamente los puntos sobre la fuente tienen valores sobre 20 dB, como era de esperarse por ser la posición igual al parlante.

El caso de los planos  $z = 25$  y  $z = 50$  son equivalentes para ver con mas detalles ir a ANEXO F.

## **6.6 INTERFAZ GRÁFICA**

Para realizar un mejor análisis de resultados del sistema de predicción y cancelación de voz, se desarrolló una interfaz gráfica desarrollada en la herramienta computacional MatLab® por su robustez matemática y sencillez de manejo permitió que la implementación del sistema se hiciera de una forma natural y amigable sin dejar de ser eficiente y sólida a nivel computacional.

El sistema desarrollado permite realizar un análisis detallado de los experimentos de predicción y cancelación de señales de voz.

Para el estudio de predicción permite la variación de parámetros de manera que se puedan analizar las técnicas para poder hacer resultados comparativos, es posible simular por las técnicas, tener la percepción visual de la voz, escucharla, cambiar el individuo, la frase o incluso agregar una nueva voz de una nueva persona para saber cómo se comportan las técnicas implementadas.

En cuanto al estudio de la cancelación de señales de voz la herramienta permite tener un bosquejo general de todas las pruebas que se efectuaron en el desarrollo del sistema de cancelación, además, permite probar el buen funcionamiento del mismo al poder realizar el experimento desde la interfaz.

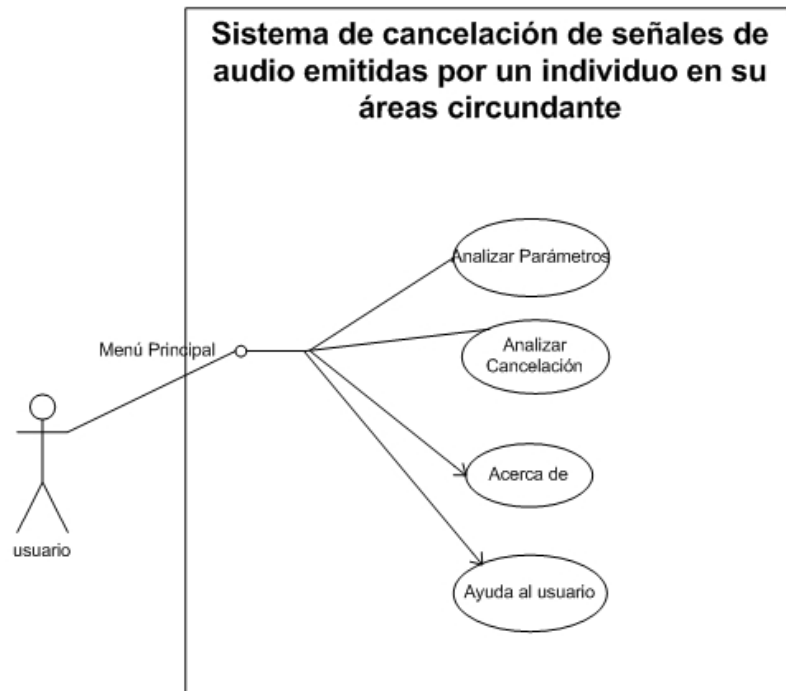
Para ilustrar la forma de interacción entre el usuario y la herramienta se presenta a continuación el caso de uso de todo el sistema.

Figura 46. Casos de uso sistema total.



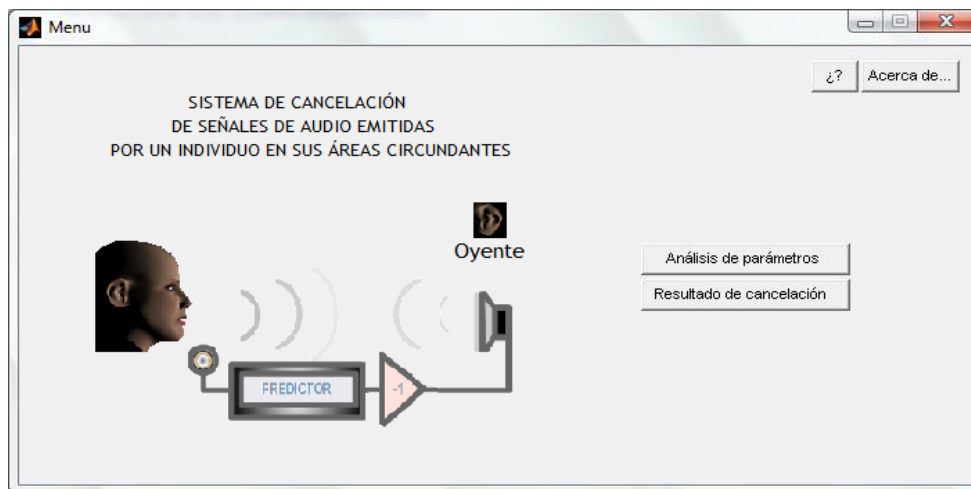
La ventana principal del sistema y de primera interacción con el usuario es el menú principal, donde la persona puede escoger que análisis desea hacer (de predicción o cancelación), consultar la ayuda al usuario y ver los créditos de la investigación como se observa en los casos de uso (figura 42)

Figura 47. Casos de uso interfaz Menú principal.



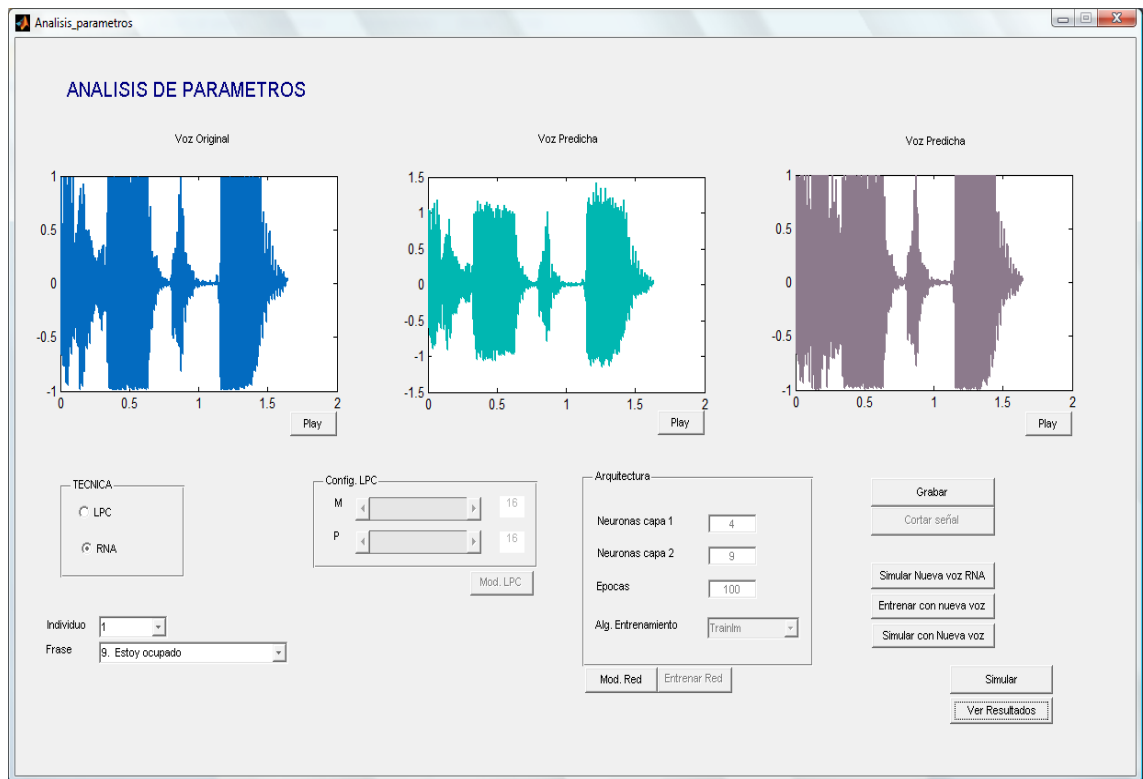
La imagen de la interfaz se visualiza en la siguiente figura

Figura 48. Ventana del Menú Principal



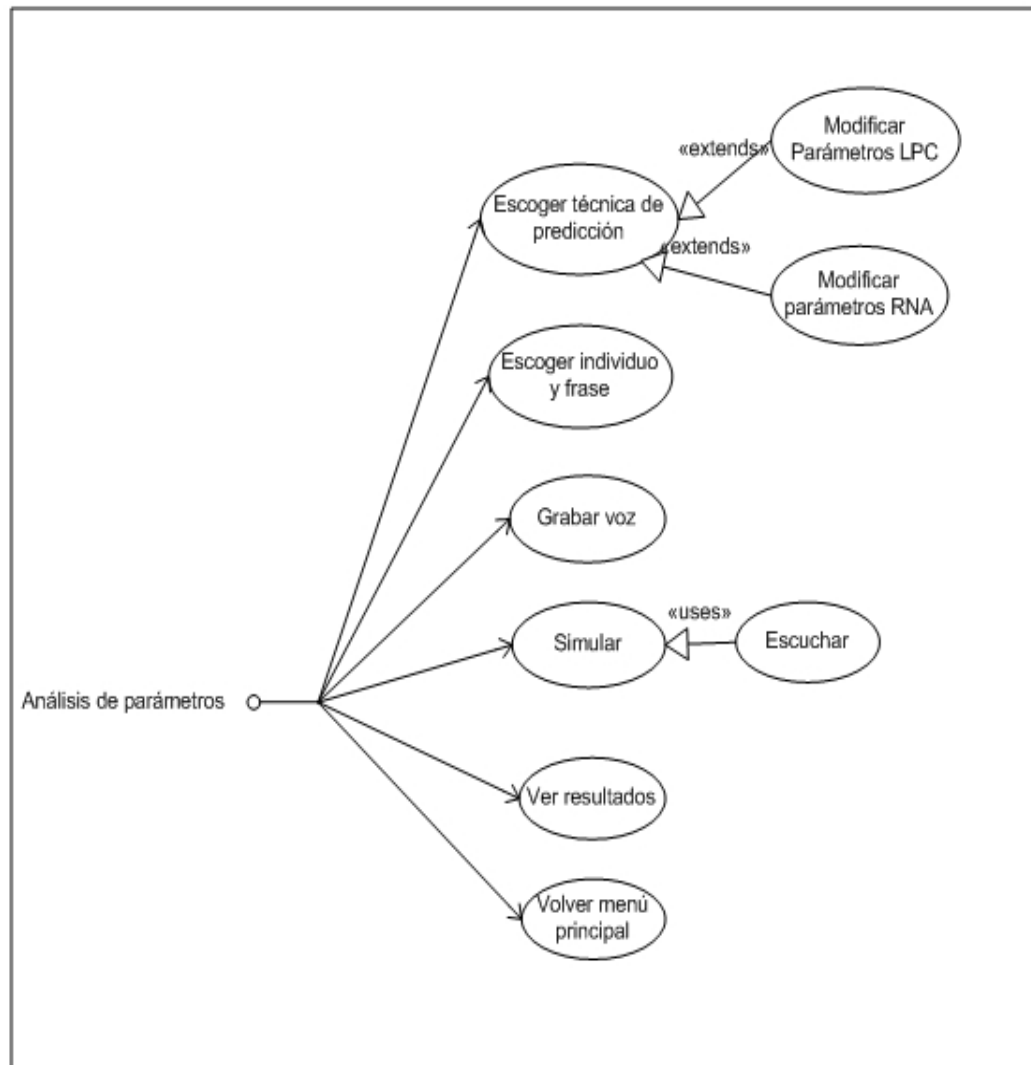
Si el usuario escoge la opción “Análisis de parámetros” se abre la ventana del mismo nombre (figura 49)

Figura 49. Ventana de análisis de Parámetros



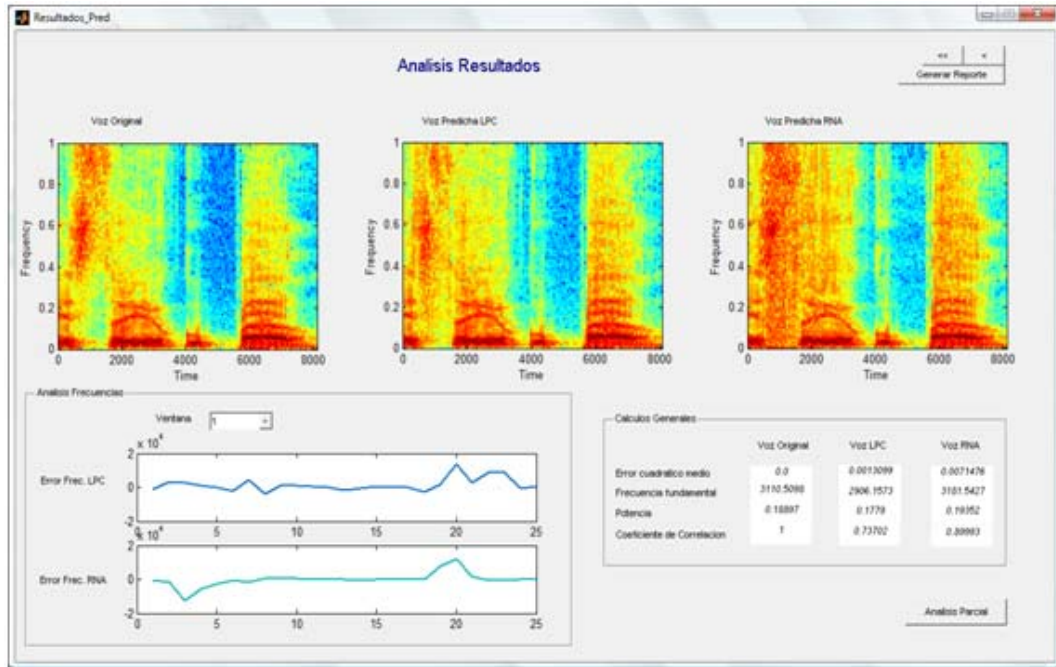
La interfaz muestra los resultados obtenidos con las configuraciones de RNA y LPC que se considera arrojan mejores resultados de predicción. Sin embargo, da la opción de variar los parámetros, permitiendo al usuario probar otros valores de p y M para la técnica de predicción lineal y verificar arquitecturas diferentes para el caso de redes neuronales. El funcionamiento de esta ventana se modela con los casos de uso que se observan la figura 50

Figura 50. Casos de uso interfaz Análisis de Parámetros.



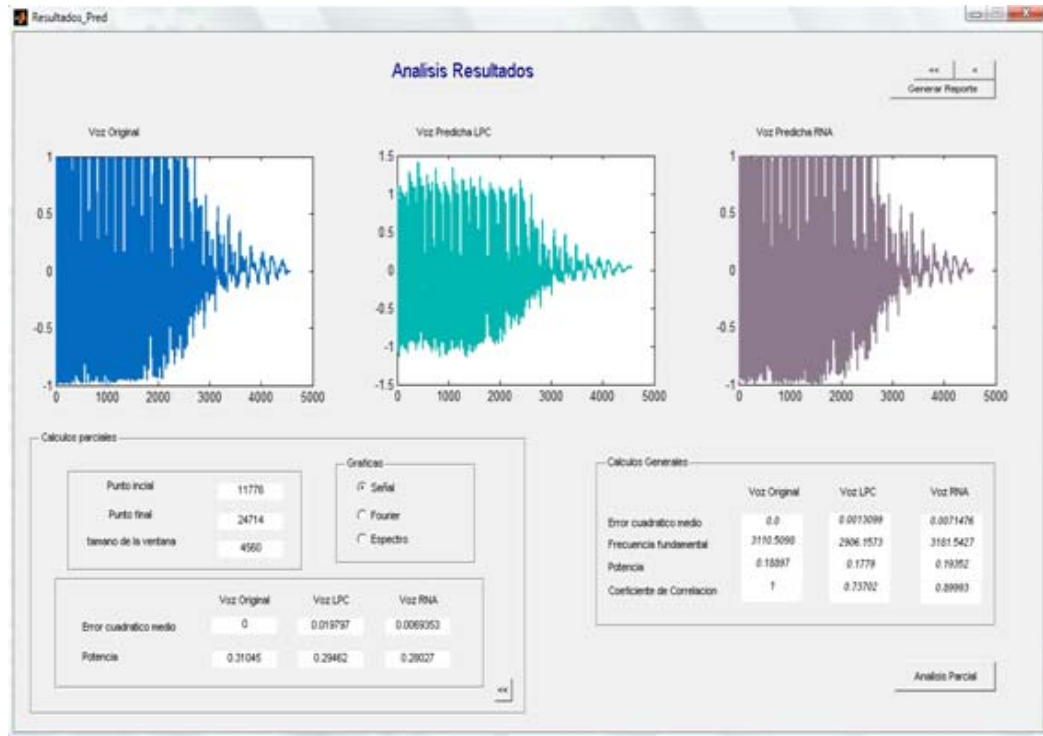
Una segunda ventana (Ver figuras 51 y 52) permite observar el comportamiento de la señal en el espacio de frecuencias por medio del espectro y a través de ventanas promedian los valores de frecuencia y analizan el error de proximidad frente al esperado. Por otro lado, se muestran los resultados matemáticos que se pueden calcular a partir de la información que se tiene de las señales predichas, dándole al usuario un número elevado de parámetros de evaluación para elegir una técnica según sus necesidades.

Figura 51. Ventana de Análisis de resultados



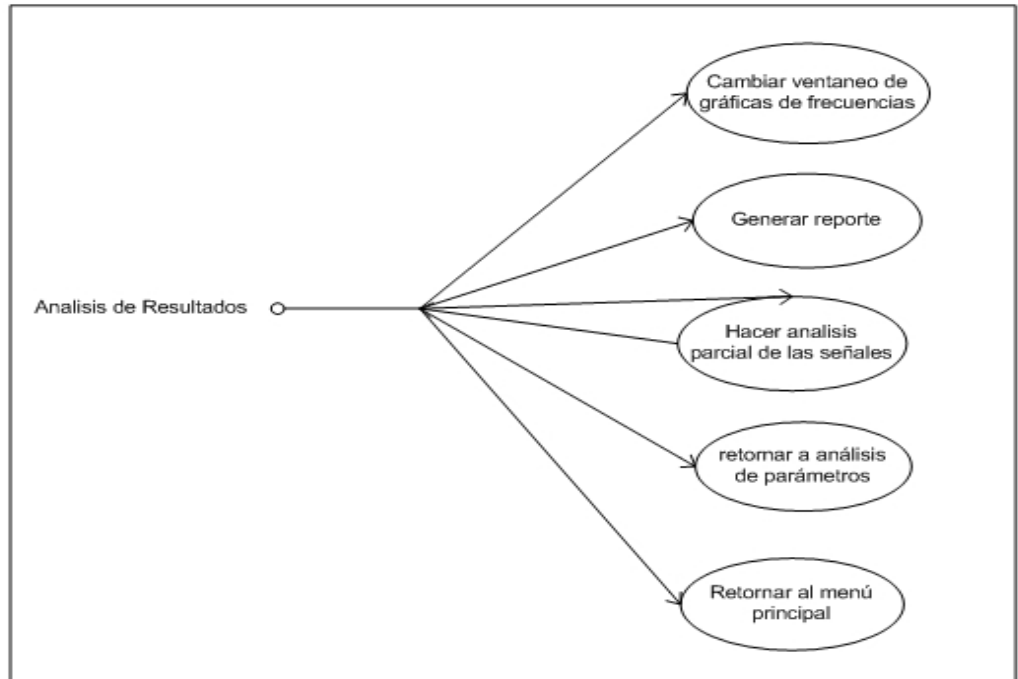
Además, es posible escoger un segmento de la muestra para hacer un análisis parcial lo cual permite ver de forma grafica y numérica en qué puntos los datos predichos tienen mayor variación con el punto de referencia. Finalmente esta ventana permite generar un reporte en Excel con los datos más relevantes de las muestras predichas.

Figura 52. Ventana de análisis de resultados para un segmento de la señal



La siguiente figura muestra los casos de usos de la ventana de análisis de resultados.

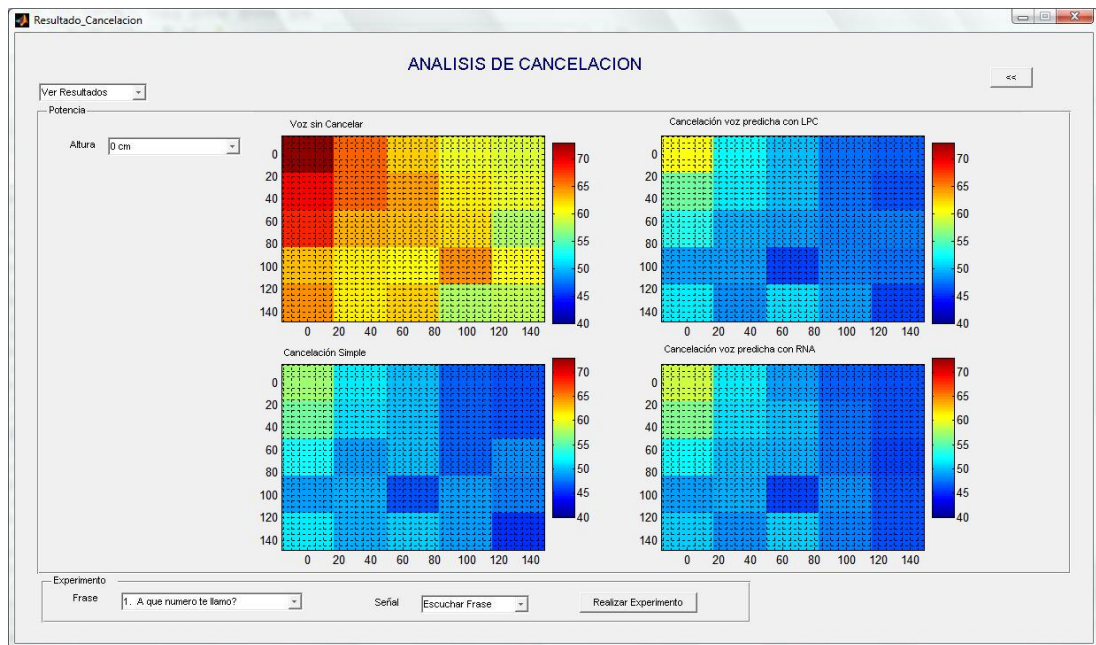
Figura 53. Casos de uso interfaz Análisis de Resultados.



Por otro lado si el usuario escoge la opción “Resultados de Cancelación” del Menú Principal es enviado a la ventana de Análisis de Cancelación en la cual el usuario puede observar todos los resultados del experimento de cancelación realizado con el sonómetro y si desea puede tomar datos realizando nuevamente el experimento desde la interfaz.

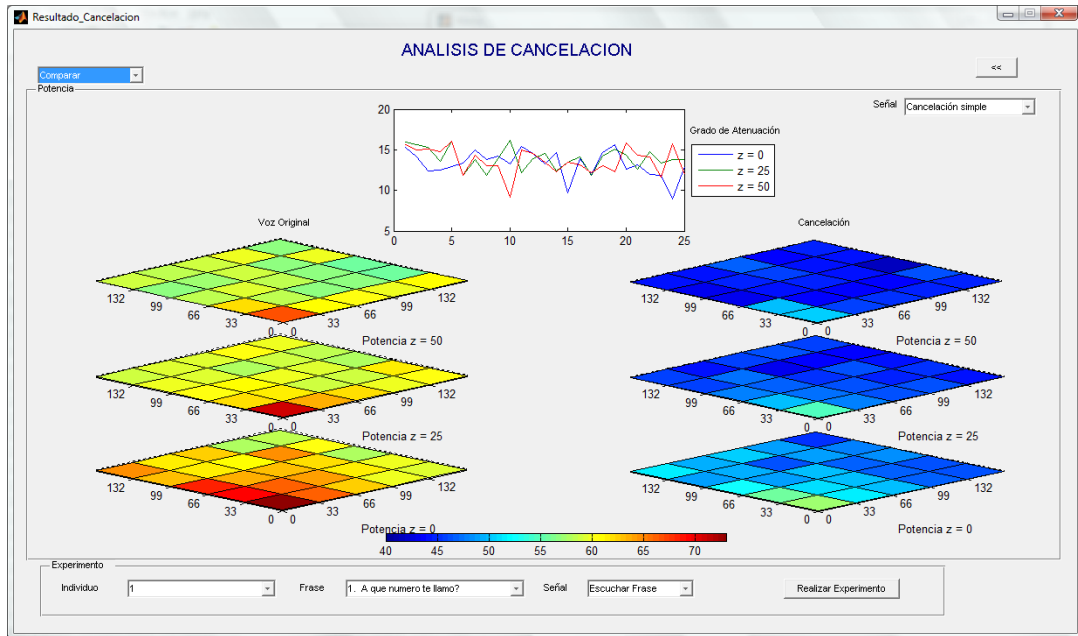
En las siguientes graficas se muestra la ventana descrita:

Figura 54. Interfaz de Resultados de Cancelación



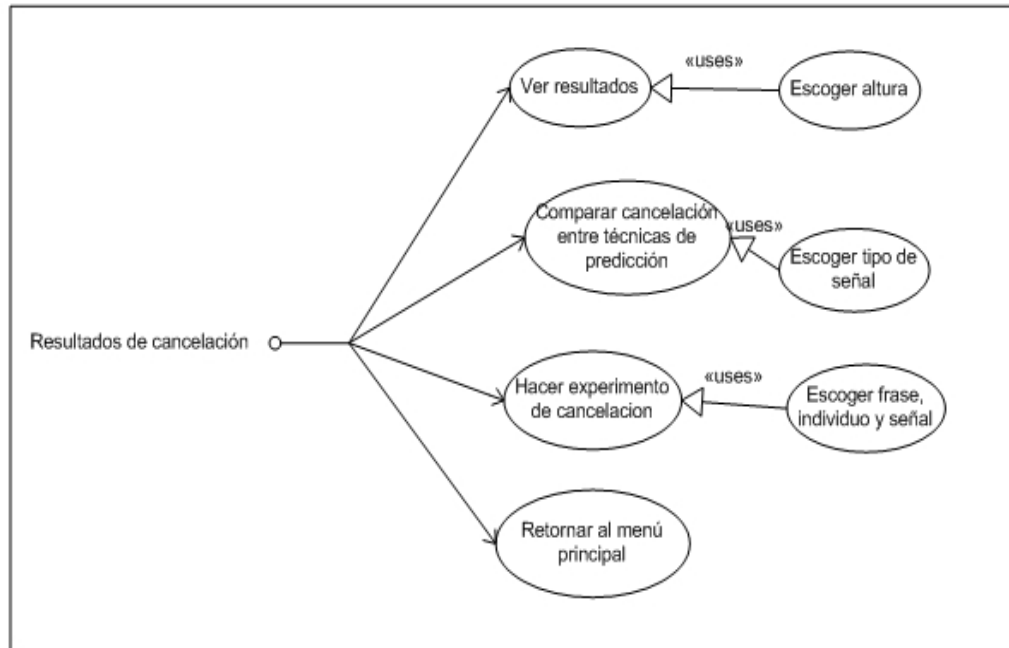
Es posible presentar al usuario una forma de análisis sencilla pero objetiva para poder realizar un análisis de potencia por plano. (Figura 55)

Figura 55. Comportamiento de la cancelación de señales de voz en el entorno



La interfaz permite ver el comportamiento de las potencias de las señales de voz en un entorno comparando cada una con las potencias de la señal sin cancelar. Se observa la resta de los valores punto a punto para ver el grado de disipación del sistema según el tipo de señal utilizada. El comportamiento de esta ventana se modela con los casos de uso de la siguiente figura:

Figura 56. Casos de uso interfaz Resultados de Cancelación



## 7. CONCLUSIONES Y TRABAJOS FUTUROS

### 7.1 CONCLUSIONES

- ▶ Es necesario tener una buena fundamentación matemática y teórica en el momento de desarrollar los algoritmos de predicción propuestos para integrar conocimiento con tecnología y así lograr resultados coherentes entre la práctica y la teoría.
- ▶ Las bases teóricas aprendidas durante la carrera, el conocimiento adquirido y la práctica alcanzada al culminar el ciclo educativo contribuyen notablemente al desarrollar un proyecto de ingeniería, en proyectos de esta magnitud es donde se logra integrar todos los campos del saber.
- ▶ La contaminación auditiva es un problema latente que por ignorancia no se percibe el impacto negativo en la comunidad. En proyectos como estos, donde es posible palpar por medio de las mediciones la problemática planteada, es preciso contribuir mediante la academia aportando soluciones a este problema.
- ▶ La base de datos debe ser coherente con la población y las frases según la investigación, de igual forma la toma de muestras de voz para trabajos relacionados con tecnologías de compresión de voces debe realizarse teniendo en cuenta las frecuencias típicas de muestreo de los dispositivos, al igual que es importante tomar una base de datos coherente y con buenas condiciones para poder realizar el trabajo.
- ▶ El método de predicción lineal es comúnmente utilizado por su simplicidad, se sacrifica calidad logrando una predicción aceptable por otra parte, el método

de predicción no lineal corresponde a sistemas más complejos y exige por tanto mayor esfuerzo computacional, a cambio se pueden variar mayor número de parámetros.

- ▶ Tanto el método lineal (LPC) como el método alternativo de predicción (RNA), tienen prestaciones favorables, son tolerantes a señales con ruido y funciona con voces en otros idiomas, la diferencia principal y fundamental para la investigación es la variación de parámetros. El método lineal tiene la limitación que sólo permite cambiar el orden de predicción, mientras que en RNA es posible realizar cambios a los parámetros de la red ó proponer nuevas alternativas en la forma de predecir las señales de voz.
- ▶ Las pruebas realizadas para los algoritmos de predicción requieren tener en cuenta la población, las frases a utilizar el ruido ambiental y la calidad de las señales de voz. Al realizar pruebas cortas con frecuencias altas y bajas, se logra percibir la diferencia, concluyendo dos cosas: primero, las frecuencias de muestreo bajas de 4KHz predicen bien pero con cierta cantidad de ruido además el tiempo de procesamiento es bajo, para frecuencias altas de 20KHz o 44KHz, se logra más nitidez en las señales, pero el tiempo de respuesta es más largo y demorado.
- ▶ El sistema implementado está diseñado para trabajar en tiempo real para lo cual se hace necesario implementarlo en trabajos posteriores en un DSP que permita hacer millones de cálculos matemáticos en milésimas de segundos, características con las que no cuenta un computador personal.
- ▶ Para realizar las pruebas de cancelación se debe tener en cuenta la presión sonora de la voz humana, las pruebas fueron tomadas a un volumen con presión sonora igual a 62 dB que corresponde a la voz normal promedio.

- ▶ El sistema de cancelación contribuye como una solución al problema de contaminación sonora planteado, disminuyendo de 10 a 14 dB aproximadamente de potencia según las condiciones del medio.
- ▶ Al tomar constante las condiciones del medio los resultados demuestran que el sistema logra disminuir el ruido al rango de un susurro lo que está catalogado como un ambiente de mucho silencio.
- ▶ La cancelación en el área circundante es posible, varía según la posición de los parlantes y la ubicación del oyente externo. Sin embargo, depende de la calidad de la señal el porcentaje de error obtenido.
- ▶ La cancelación utilizando señales por las dos técnicas de predicción es posible. Teniendo mejores resultados con RNA debido a su carácter no lineal que hace que la señal predicha tenga características más próximas a la señal original que la obtenida con LPC.
- ▶ El texto elaborado presenta una síntesis y una interpretación de resultados que sirve de apoyo para la elaboración de otros proyectos, que requieran de la implementación de técnicas de predicción de señales de voz, debido a que, se muestra de manera sencilla el desarrollo y prueba de los algoritmos.
- ▶ La formación obtenida por los autores en el desarrollo del presente proyecto otorgó lecciones importantes para la vida profesional y personal, tales como: el manejo del tiempo, el trabajo en grupo, la organización de actividades a realizar, las relaciones humanas, la forma de gestionar permisos y recursos, entre otras son enseñanzas importantes para trabajos futuros.

## **7.2. TRABAJOS FUTUROS**

Se propone retomar este trabajo y mejorar su efectividad para predecir mayor número de muestras en cada iteración, mejorar el tiempo de simulación al predecir las señales de voz, para esto es recomendable implementar otras técnicas de predicción como las descritas en el estado del arte.

Igualmente se plantea implementar los algoritmos propuestos, integrados en un Procesador Digital de Señales (DSP), para lo cual se plantea un diseño simple (ver ANEXO G) en cual sea posible obtener el sistema de cancelación en tiempo real y así comprobar la efectividad de la innovación con esta tecnología.

El campo de aplicación e investigación de las señales de voz requiere mayor estudio, por tal motivo, se propone realizar un estudio detallado de propiedades inherentes a las señales de voz, que permitan plantear nuevos métodos en la predicción y cancelación de señales de voz.

## BIBLIOGRAFÍA

- [1] AFFES, S. GRENIER, Y. A source subspace tracking array of microphones for double talk situations. Atlanta. En: Acoustics, Speech, and Signal Processing, IEEE (may. 1996); p. 909-912
- [2] ALONSO Marcelo y FINN Edward. Física. Washington: Fondo Educativo Interamericano, 1999.
- [3] ATAL, B.S. The history of linear prediction. En: Signal Processing Magazine, IEEE. (mar. 2006); p.154 – 161
- [4] FAUNDEZ, Marcos, BAYÉS, Vallverdú y MONTE, Enric. Modelado predictivo no lineal de la señal de voz aplicado a codificación y reconocimiento del locutor. Tesis Doctoral. Universidad Politécnica de Cataluña, Departamento de teoría de señal y comunicación. 1998.
- [5] FIGUEIREDO, Rui y AKAY, Enis. Neural speech predictor. Bucharest En: Circuits and Systems for Communications, IEEE (jul. 2008); p. 47 - 48
- [6] GOLD, Ben y MORGAN, Nelson. Speech and audio signal processing: Processing and perception of speech and music. New York: John Wiley & Sons, 1999.
- [7] GIOWACZ, Adam. y GOWACZ, Witold. Sound recognition of DC machine with application of LPC and metrics. Krakow. En: Human System Interactions, 2008 Conference, IEEE. (may 2008); p. 322 – 327

- [8] HACIOGLU, K. y TURKYILMAZ, R. On the use of radial basis function networks for nonlinear speech processing. Tel-Aviv, Israel. En: Electrotechnical Conference, IEEE (may. 1998); p. 29 - 33
- [9] HAYKIN, Simon. Neural networks. A comprehensive foundation. Segunda edición. Montreal: Prentice-Hall, 1999.
- [10] KONDO, Kazuhiro y NAKAGAWA, Kiyoshi. On Long Term Prediction for Active Cellular Speech Emission Control. En: Proceedings of the Internoise. Korea. (aug. 2003); p. 26-15
- [11] LARIVIERE, Jeff. y GOUBRAN, Rafii. Noise-reduced GMDF for acoustic echo cancellation and speech recognition in mobile environments. Boston. En: Vehicular Technology Conference, IEEE 2000; p. 2969-2972
- [12] MESSERSCHMITT, David. Echo Cancellation in Speech and Data. California. En: Selected Areas in Communications, IEEE (mar. 1984); p. 283- 297
- [13] RABINER, Lawrence y SCHAFER, Ronald. Digital processing of speech signals. New Jersey: Prentice Hall, 1979.
- [14] SALAZAR, Jordi. Procesadores digitales de señales (DPS) Arquitectura y criterios de selección. Departamento ingeniería electrónica. Centro de sistemas y sensores eléctricos. Universidad Politécnica de Cataluña. 2000.

[15] SIMONITE Tom, Nerve-tapping neckband used in 'telepathic' chat. En: New Scientist Tech (<http://technology.newscientist.com/article/dn13449-nervetapping-neckband-allows-telepathic-chat.html>Marzo%202008) Ultima visita: 15 de octubre de 2008

[16] STAVRAKOUDIS, Dimitris. THEOCHARIS, John. Theocharis pipelined recurrent fuzzy neural networks for nonlinear adaptive speech prediction. Thessaloniki. En: Systems, Man, and Cybernetics, IEEE (oct. 2007), p. 1305 – 1320

[17] TRABELSI, Boyer. SAVARIA, Boukadoum, Improving LPC analysis of speech in additive noise. Montreal. En: Circuits and Systems for Communications, IEEE (agosto 2007); p. 93 – 96

[18] WANG, An-Hong. SUN Zhi-Yi. LI, Zhi-Hong, et al. A dynamic wavelet neural network for prediction of speech signal. Taiyuan. En: Machine Learning and Cybernetics, IEEE (agosto 2004); p. 3524- 3528

[19] ZHANG, Rubo. TIAN, Wu. XUEYAO, Li et al. Speech stream detection in strong noise based on linear prediction. Singapore. En: Industrial Electronics and Applications, IEEE. (mayo 2006); p. 1 - 4

[20] ZHANG, Xueying. y GUO, Yueling. HOU, Xuemei. A speech recognition method of isolated words based on modified LPC cepstrum. Fremont. En: Granular Computing, IEEE. (nov. 2007); p. 481 - 481

**ANEXO A**  
**PRUEBAS DE ALGORITMOS DE ENTRENAMIENTO**

Pruebas de los algoritmos de entrenamiento variando el número de épocas, para la arquitectura: 4 neuronas en la primera capa oculta, 9 neuronas en la segunda capa oculta y 1 neurona de salida

<i>Algoritmo de entrenamiento</i>	<i>Épocas</i>	<i>Función de Rendimiento</i>
<b>Traingdx</b> – Variable Learning Rate Backpropagation	100	9,71E-04
	200	6,81E-04
	300	6,06E-04
	400	5,59E-04
	500	5,27E-04
	600	4,98E-04
	700	4,79E-04
	800	4,65E-04
	900	4,54E-04
	1000	4,43E-04
	1500	1,85E-04
2000	1,60E-04	
<b>Trainlm</b> – Levenberg – Marquardt	100	1,84E-06
	200	1,83E-06
	300	1,83E-06
	400	1,82E-06
	500	1,80E-06
	600	1,79E-06
	700	1,79E-06
	800	1,79E-06
	900	1,78E-06
	1000	1,78E-06
<b>Trainbfg</b> – BFGS Quasi – Newton	100	1,00E-04
	200	5,82E-05
	300	4,16E-05
	400	2,10E-05
	500	1,80E-05
	600	1,25E-05
	700	1,17E-05
	800	1,15E-05

	900	1,07E-05
	1000	1,04E-05
<b>Trainscg</b> – Scaled Conjugate Gradient	100	0,92
	200	0,20
	300	0,17
	400	0,09
	500	0,07
	600	0,06
	700	0,05
	800	0,05
	900	0,05
	1000	0,04
<b>Trainscgf</b> – Fletcher-Powell Conjugate Gradient	100	5,53E-05
	200	2,10E-05
<b>Trainoss</b> – One-Step secant	100	3,19E-04
	200	2,52E-04
	300	1,58E-04
	400	1,26E-04
	500	7,91E-05
	600	7,25E-05
	700	4,71E-05
	800	4,56E-05
	900	3,84E-05
	1000	3,70E-05

## ANEXO B

### PRUEBAS DE CANCELACIÓN ESCENARIO 1 CENTIC Ruido ambiental 47 dB

x	y	Voz normal Z=0cm	Cancelación. Optima Z= 0cm	Voz normal Z=25cm	Cancelación. Optima Z= 25cm	Voz normal Z= 50 cm	Cancelación. Optima Z= 50 cm	Potencia disipada Z=0cm	Potencia Disipada Z=25cm	Potencia Disipada Z=50cm	
0	0	70,3	60	65,5	52,1	57,5	48,5	10,3	13,4	9	
0	50	62	52,7	58,3	48,7	52,8	46,6	9,3	9,6	6,2	
0	100	61	49,3	56,7	48,9	50,3	46,7	11,7	7,8	3,6	
0	150	58,1	48	52,7	46,3	50,2	46,5	10,1	6,4	3,7	
0	200	56	45,8	50,9	46,2	50,3	46,5	10,2	4,7	3,8	
50	0	64,1	56,7	53,2	48,1	56,5	47,4	7,4	5,1	9,1	
50	50	61,1	49,9	52,2	47,8	51,4	46,4	11,2	4,4	5	
50	100	61,9	46,8	52,7	46,6	50,3	46,5	15,1	6,1	3,8	
50	150	59,6	47,6	53,6	45,7	50,2	46,3	12	7,9	3,9	
50	200	57,1	46,2	49,6	46,4	50,1	46,6	10,9	3,2	3,5	
100	0	59,7	55,3	53,9	47	55,9	47,4	4,4	6,9	8,5	
100	50	59,4	48,8	54,6	45,9	50,2	46,6	10,6	8,7	3,6	
100	100	60,6	50	59,7	49,8	50,3	46,5	10,6	9,9	3,8	
100	150	60,7	50,4	56,4	47,3	50,3	46,4	10,3	9,1	3,9	
100	200	58,2	48,6	54,1	48	50,1	46,4	9,6	6,1	3,7	
150	0	60,2	48,9	55	45,6	55,2	47,2	11,3	9,4	8	
150	50	57,4	47	55,8	45	50,4	46,6	10,4	10,8	3,8	
150	100	58,2	48,7	54,4	48	50,2	46,4	9,5	6,4	3,8	
150	150	57,2	46,9	55,8	54,5	50,2	46,5	10,3	1,3	3,7	
150	200	58,7	47,3	53,5	47,6	50,1	46,4	11,4	5,9	3,7	
200	0	58,7	49,1	52,9	45,8	55,2	47,2	9,6	7,1	8	
200	50	58,3	47,9	55,2	45,8	50,2	46,7	10,4	9,4	3,5	
200	100	56,9	47,3	50,8	46,8	50,3	46	9,6	4	4,3	
200	150	56,1	46,6	51	44,5	50,1	46,2	9,5	6,5	3,9	
200	200	57,2	47,3	51,7	47,2	50,1	46,4	9,9	4,5	3,7	
					<b>promedio de potencia disipada</b>				10,224	6,984	4,86

**PRUEBAS DE CANCELACIÓN ESCENARIO 2**  
**CTI**  
**Ruido ambiental 33 dB**

X	y	Voz normal Z=0cm	Cancelación. Optima Z= 0cm	Voz normal Z=25cm	Cancelación. Optima Z= 25cm	Voz normal Z= 50 cm	Cancelación. Optima Z= 50 cm	Potencia disipada Z=0cm	Potencia Disipada Z=25cm	Potencia Disipada Z=50cm	
0	0	72,5	57,2	70,5	54,6	66,4	50,8	15,3	15,9	15,6	
0	33	65,8	51,7	63,7	48,1	60,8	45,8	14,1	15,6	15	
0	66	62,3	49,9	61,8	46,5	60,6	45,5	12,4	15,3	15,1	
0	99	59,6	47,1	60	46,4	59,6	44,8	12,5	13,6	14,8	
0	132	59,5	46,6	60,5	44,4	60,9	44,8	12,9	16,1	16,1	
33	0	68,9	55,5	62	50,1	61,8	50	13,4	11,9	11,8	
33	33	66	51	61,2	47,4	58,3	44	15	13,8	14,3	
33	66	64	50,2	58,6	46,7	57	44	13,8	11,9	13	
33	99	61,4	47,2	58,8	44,9	56,7	43,7	14,2	13,9	13	
33	132	59,7	46,4	61,5	45,3	55,6	46,4	13,3	16,2	9,2	
66	0	68,2	52,8	60,2	48	58,5	43,5	15,4	12,2	15	
66	33	63,6	49,1	60,9	47	59,1	44,5	14,5	13,9	14,6	
66	66	63,1	49,8	59,8	45,2	58,5	45	13,3	14,6	13,5	
66	99	61,9	47,2	59,1	46,7	56,9	44,6	14,7	12,4	12,3	
66	132	58	48,3	57,8	44,3	55,5	42	9,7	13,5	13,5	
99	0	63	49,1	59,9	45,8	56,7	43,5	13,9	14,1	13,2	
99	33	61,6	49,6	59,4	47,5	57,1	44,9	12	11,9	12,2	
99	66	61,1	46,4	57,8	43,6	59	46	14,7	14,2	13	
99	99	64,4	48,8	58,9	43,8	56,7	44,4	15,6	15,1	12,3	
99	132	60,5	47,9	58,3	44	60,1	44,3	12,6	14,3	15,8	
132	0	64,6	51,4	58,6	46	58,6	44,3	13,2	12,6	14,3	
132	33	61,5	49,5	59,1	44,3	58,4	44,3	12	14,8	14,1	
132	66	62,2	50,4	59,4	46	58,6	47	11,8	13,4	11,6	
132	99	57,7	48,8	60,3	46,5	59,6	43,9	8,9	13,8	15,7	
132	132	58,4	45,5	59,7	45,9	57	45	12,9	13,8	12	
					<b>promedio de potencia disipada</b>			10,224	6,984	4,86	

**ANEXO C**  
**POTENCIA DE CANCELACIÓN CTI SIN RUIDO AMBIENTAL DE 33dB**

x	y	Voz normal (Z=0)	cancelación simple (Z=0)	Voz normal (Z=25 cm)	Cancelación (Z=25 cm)	Voz normal (Z=50 cm)	Cancelación (Z=50 cm)
0	0	39,5	24,2	37,5	21,6	33,4	17,8
0	33	32,8	18,7	30,7	15,1	27,8	12,8
0	66	29,3	16,9	28,8	13,5	27,6	12,5
0	99	26,6	14,1	27	13,4	26,6	11,8
0	132	26,5	13,6	27,5	11,4	27,9	11,8
33	0	35,9	22,5	29	17,1	28,8	17
33	33	33	18	28,2	14,4	25,3	11
33	66	31	17,2	25,6	13,7	24	11
33	99	28,4	14,2	25,8	11,9	23,7	10,7
33	132	26,7	13,4	28,5	12,3	22,6	13,4
66	0	35,2	19,8	27,2	15	25,5	10,5
66	33	30,6	16,1	27,9	14	26,1	11,5
66	66	30,1	16,8	26,8	12,2	25,5	12
66	99	28,9	14,2	26,1	13,7	23,9	11,6
66	132	25	15,3	24,8	11,3	22,5	9
99	0	30	16,1	26,9	12,8	23,7	10,5
99	33	28,6	16,6	26,4	14,5	24,1	11,9
99	66	28,1	13,4	24,8	10,6	26	13
99	99	31,4	15,8	25,9	10,8	23,7	11,4
99	132	27,5	14,9	25,3	11	27,1	11,3
132	0	31,6	18,4	25,6	13	25,6	11,3
132	33	28,5	16,5	26,1	11,3	25,4	11,3
132	66	29,2	17,4	26,4	13	25,6	14
132	99	24,7	15,8	27,3	13,5	26,6	10,9
132	132	25,4	12,5	26,7	12,9	24	12

**ANEXO D**  
**PRUEBAS CUALITATIVAS CON USUARIO FINAL**  
**ENCUESTA**

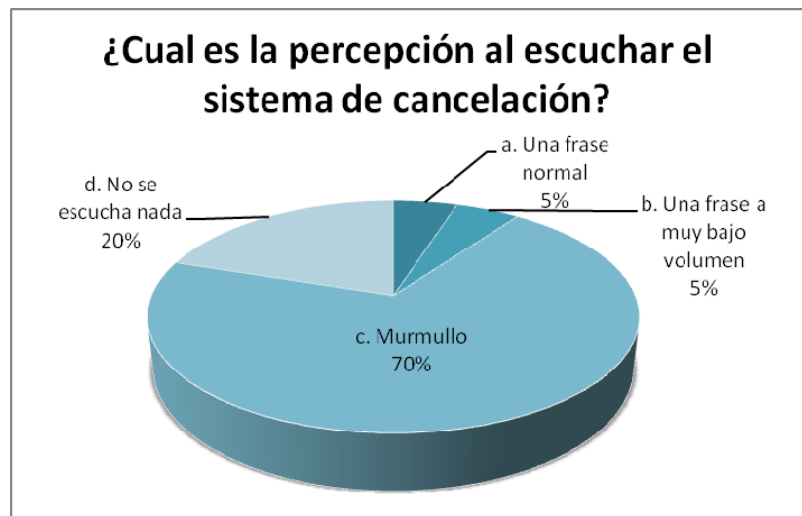
A cada una de las personas encuestadas después de poner en funcionamiento el sistema de cancelación se les formularon las siguientes preguntas:

1. ¿Que frase escucha?
- 2.Cuál es la percepción al escuchar el sistema de cancelación:
  - a. Una frase normal
  - b. Una frase a muy bajo volumen
  - c. Un murmullo
  - d. No se escucha nada
3. Identifica el género de la persona parlante  
\_\_\_ Si        \_\_\_ No
4. El sonido que alcanza a percibir es:
  - a. Molesto
  - b. No se alcanza a percibir
  - c. Es indiferente
  - d. No sabe / no responde
5. Encuentra alguna diferencia entre la intensidad del sonido normal y el sonido con el sistema de cancelación:
  - a. Tienen el mismo volumen
  - b. El sonido normal tiene menor volumen
  - c. El sistema de cancelación atenúa la señal
  - d. En ninguno de los casos se logra identificar la frase.

## Resultados de la encuesta

1. De cada 10 frases que escuchaba la persona solo 2 de los 20 individuos entendieron algunas palabras correspondientes a las frases, en un caso particular la persona entendió una frase completamente diferente a la real.

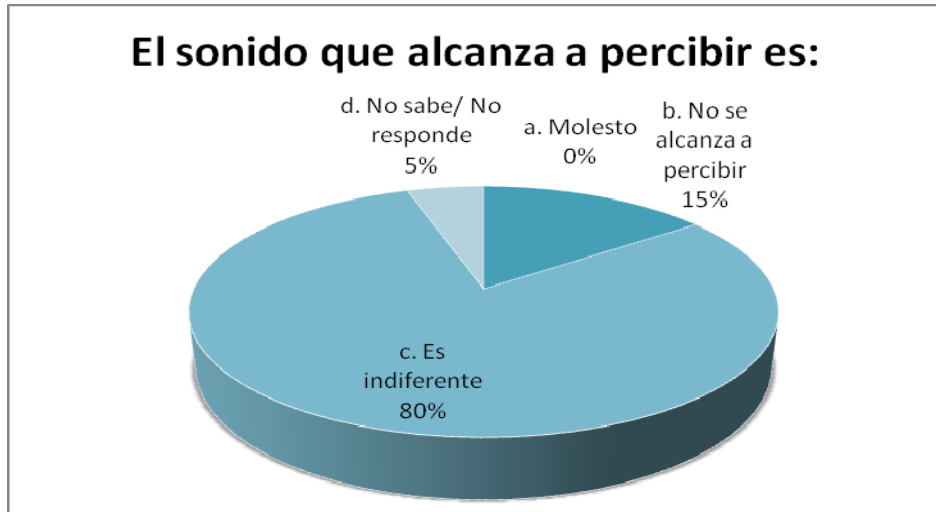
2. El 5% de los encuestados escucho una frase normal, 5% una frase a muy bajo volumen, 70 % un murmullo y el 20% definitivamente no escucha nada.



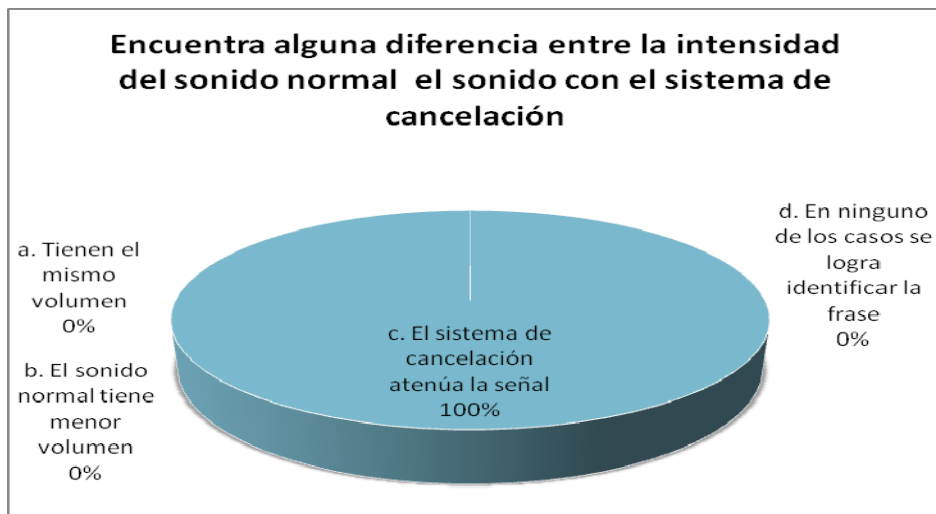
3. El 80 % identifica el género de la persona y el 20% no debido a que no escuchan nada.



4. El 15% no alcanza a percibir nada, para el 80% de las personas es indiferente y el 5% no sabe no responde.



5. El 100% está de acuerdo con que el sistema de cancelación atenúa la señal.



La siguiente tabla muestra los resultados de la encuesta por pregunta.

<i>Individuo</i>	<i>Pregunta 2</i>	<i>Pregunta 3</i>	<i>Pregunta 4</i>	<i>Pregunta 5</i>
1	c	Si	C	C
2	b	Si	C	C
3	c	Si	C	C
4	c	Si	C	C
5	c	Si	C	C
6	c	Si	C	C
7	c	Si	C	C
8	d	No	B	C
9	d	No	B	C
10	a	Si	C	C
11	c	Si	C	C
12	c	Si	C	C
13	c	Si	C	C
14	d	No	D	C
15	c	Si	C	C
16	c	Si	C	C
17	c	Si	C	C
18	d	No	B	C
19	c	Si	C	C
20	c	Si	C	C

**ANEXO E**  
**CANCELACIÓN CON TÉCNICAS DE PREDICCIÓN (CTI)**

**Z=0 m**

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	72,5	57,2	60,6	58,6
0	33	65,8	51,7	52	51,7
0	66	62,3	49,9	49,9	49,1
0	99	59,6	47,1	47,6	47,1
0	132	59,5	46,6	46,8	46,3
33	0	68,9	55,5	55,6	56,3
33	33	66	51	51,6	51,2
33	66	64	50,2	50,2	50
33	99	61,4	47,2	47,3	47,6
33	132	59,7	46,4	46,6	46,2
66	0	68,2	52,8	53,4	52
66	33	63,6	49,1	49,3	49,9
66	66	63,1	49,8	48,8	49,4
66	99	61,9	47,2	47,9	47,3
66	132	58	48,3	47,8	46,1
99	0	63	49,1	49,2	48,8
99	33	61,6	49,6	49	49,4
99	66	61,1	46,4	46,1	46,1
99	99	64,4	48,8	48,2	48,3
99	132	60,5	47,9	47,5	46,3
132	0	64,6	51,4	51,7	50,8
132	33	61,5	49,5	48,3	48,7
132	66	62,2	50,4	51,1	50,5
132	99	57,7	48,8	48,8	48,1
132	132	58,4	45,5	45,8	46,2

**Z= 25cm**

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	70,5	54,6	53,7	54,8
0	33	63,7	48,1	48,5	48

0	66	61,8	46,5	46,9	46,5
0	99	60	46,4	46,8	46,3
0	132	60,5	44,4	45,6	44,9
33	0	62	50,1	50,7	51,3
33	33	61,2	47,4	47,6	47,6
33	66	58,6	46,7	47,2	47
33	99	58,8	44,9	46	45
33	132	61,5	45,3	47	46,3
66	0	60,2	48	48,8	48,8
66	33	60,9	47	47,8	46,9
66	66	59,8	45,2	46,8	45,1
66	99	59,1	46,7	46,8	46,5
66	132	57,8	44,3	46,3	45,5
99	0	59,9	45,8	47	45,9
99	33	59,4	47,5	48	48
99	66	57,8	43,6	44,1	43,4
99	99	58,9	43,8	44,5	43,6
99	132	58,3	44	44,6	44
132	0	58,6	46	46,4	46,1
132	33	59,1	44,3	45	45
132	66	59,4	46	45,7	46
132	99	25	60,3	46,5	46,6
132	132	25	59,7	45,9	46,2

Z=50 cm

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	66,4	50,8	51	50,9
0	33	60,8	45,8	46	45,8
0	66	60,6	45,5	45,7	45,8
0	99	59,6	44,8	45,3	44,9
0	132	60,9	44,8	45,2	44,9
33	0	61,8	50	50,7	50,8
33	33	58,3	44	45,8	45,6
33	66	57	44	44,4	44,2

33	99	56,7	43,7	43,7	43,9
33	132	55,6	46,4	46,7	46,9
66	0	58,5	43,5	44,7	44,4
66	33	59,1	44,5	45	45
66	66	58,5	45	44,9	45
66	99	56,9	44,6	44,4	44,8
66	132	55,5	42	42,9	42,6
99	0	56,7	43,5	44,6	44,4
99	33	57,1	44,9	44,5	44,4
99	66	59	46	46,5	46,3
99	99	56,7	44,4	44,6	44,4
99	132	60,1	44,3	44,3	44,3
132	0	58,6	44,3	44,6	44,2
132	33	58,4	44,3	44,6	44,4
132	66	58,6	47	47,2	47
132	99	59,6	43,9	44,7	44,5
132	132	57	45	45,6	45,6

**ANEXO F**  
**POTENCIA DE CANCELACIÓN CON TÉCNICAS SIN RUIDO AMBIENTAL DE 33dB**

Z=0

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	39,5	24,2	27,6	25,6
0	33	32,8	18,7	19	18,7
0	66	29,3	16,9	16,9	16,1
0	99	26,6	14,1	14,6	14,1
0	132	26,5	13,6	13,8	13,3
33	0	35,9	22,5	22,6	23,3
33	33	33	18	18,6	18,2
33	66	31	17,2	17,2	17
33	99	28,4	14,2	14,3	14,6
33	132	26,7	13,4	13,6	13,2
66	0	35,2	19,8	20,4	19
66	33	30,6	16,1	16,3	16,9
66	66	30,1	16,8	15,8	16,4
66	99	28,9	14,2	14,9	14,3
66	132	25	15,3	14,8	13,1
99	0	30	16,1	16,2	15,8
99	33	28,6	16,6	16	16,4
99	66	28,1	13,4	13,1	13,1
99	99	31,4	15,8	15,2	15,3
99	132	27,5	14,9	14,5	13,3
132	0	31,6	18,4	18,7	17,8
132	33	28,5	16,5	15,3	15,7
132	66	29,2	17,4	18,1	17,5
132	99	24,7	15,8	15,8	15,1
132	132	25,4	12,5	12,8	13,2

Z=25 cm

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	37,5	21,6	20,7	21,8
0	33	30,7	15,1	15,5	15
0	66	28,8	13,5	13,9	13,5
0	99	27	13,4	13,8	13,3
0	132	27,5	11,4	12,6	11,9
33	0	29	17,1	17,7	18,3
33	33	28,2	14,4	14,6	14,6
33	66	25,6	13,7	14,2	14
33	99	25,8	11,9	13	12
33	132	28,5	12,3	14	13,3
66	0	27,2	15	15,8	15,8
66	33	27,9	14	14,8	13,9
66	66	26,8	12,2	13,8	12,1
66	99	26,1	13,7	13,8	13,5
66	132	24,8	11,3	13,3	12,5
99	0	26,9	12,8	14	12,9
99	33	26,4	14,5	15	15
99	66	24,8	10,6	11,1	10,4
99	99	25,9	10,8	11,5	10,6
99	132	25,3	11	11,6	11
132	0	25,6	13	13,4	13,1
132	33	26,1	11,3	12	12
132	66	26,4	13	12,7	13
132	99	-8	27,3	13,5	13,6
132	132	-8	26,7	12,9	13,2

Z=50 cm

<i>Posición en X</i>	<i>Posición en y</i>	<i>Voz sin Cancelar</i>	<i>Cancelación Simple</i>	<i>Cancelación_Pred. LPC</i>	<i>Cancelación_Pred. RNA</i>
0	0	33,4	17,8	18	17,9
0	33	27,8	12,8	13	12,8
0	66	27,6	12,5	12,7	12,8
0	99	26,6	11,8	12,3	11,9

0	132	27,9	11,8	12,2	11,9
33	0	28,8	17	17,7	17,8
33	33	25,3	11	12,8	12,6
33	66	24	11	11,4	11,2
33	99	23,7	10,7	10,7	10,9
33	132	22,6	13,4	13,7	13,9
66	0	25,5	10,5	11,7	11,4
66	33	26,1	11,5	12	12
66	66	25,5	12	11,9	12
66	99	23,9	11,6	11,4	11,8
66	132	22,5	9	9,9	9,6
99	0	23,7	10,5	11,6	11,4
99	33	24,1	11,9	11,5	11,4
99	66	26	13	13,5	13,3
99	99	23,7	11,4	11,6	11,4
99	132	27,1	11,3	11,3	11,3
132	0	25,6	11,3	11,6	11,2
132	33	25,4	11,3	11,6	11,4
132	66	25,6	14	14,2	14
132	99	26,6	10,9	11,7	11,5
132	132	24	12	12,6	12,6

## **ANEXO G**

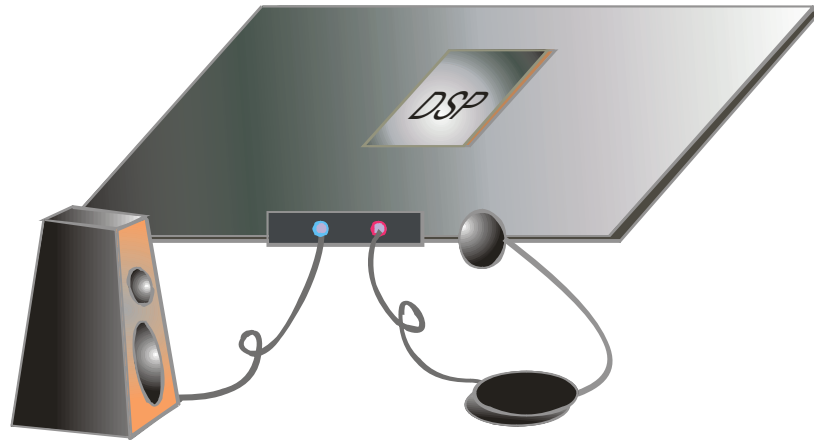
### **TRABAJOS FUTURO: DSP**

Los DSP son microprocesadores de propósitos específicos, especialmente diseñados para el procesamiento digital de señales. Cuenta con grandes ventajas frente al procesador de un computador tradicional como lo son la velocidad, la organización de memoria, la simplicidad en el formato aritmético o la arquitectura interna, que hacen que sean o no adecuados para una aplicación en particular.

Estos sistemas realizan operaciones matemáticas a señales digitalizadas necesarias para el procesamiento en tiempo real por que poseen una arquitectura especializada para acelerar los cálculos matemáticos. Muy comúnmente cuentan con arquitecturas de memoria para tener acceso múltiple que permita la carga de varios operadores simultáneamente. Cuenta con una variedad de modos especiales de direccionamiento y características de control de flujo de programas diseñados para acelerar la ejecución de operaciones repetitivas. Además, con un DSP la comunicación con los demás componentes del sistema se realiza a mayor velocidad por que incluyen periféricos e interfaces de entrada/salida que aumentan la velocidad por lo cual es ideal para trabajar con señales emulando un dispositivo móvil.

La siguiente fase para proyectos futuros, continuando con la labor de disminuir el problema planteado, consiste en implementar los algoritmos estudiados en un sistema DSP que cuente con las características necesarias para el buen funcionamiento en tiempo real. El montaje planteado es el ilustrado en la figura siguiente, consiste básicamente en un micrófono, un parlante y el sistema del DSP.

Figura 1: Sistema DSP planteado



Se plantea que las señales de entrada se convierten en una secuencia de muestras después de ser capturadas por un traductor (micrófono) y transformadas por un convertidor análogo – digital. El núcleo de este sistema es un DSP que solventa la necesidad de la carga computacional intensa.

Este montaje puede ser probado en tiempo real, la persona habla frente al micrófono que captura la señal que posteriormente va a ser procesada por el DSP, el cual envía información de respuesta al parlante, este proceso se realiza en milésimas de segundos de manera que responde casi automáticamente haciendo posible la predicción y posterior atenuación de la señal emitida por el individuo. Una vez probado el funcionamiento en el DPS, el trabajo podrá ser implementado en un dispositivo móvil que cuente con las características necesarias.

## ANEXO H

### PREDICCIÓN DE SEÑALES DE VOZ MEDIANTE COEFICIENTES DE PREDICCIÓN LINEAL Y REDES NEURONALES ARTIFICIALES

#### JULY PAOLA LICONA BLANCO

Estudiante Ingeniería de Sistemas e Informática  
Universidad Industrial de Santander  
[jplicona@sis25.uis.edu.co](mailto:jplicona@sis25.uis.edu.co)

#### ADRIANA ROCIO GARCIA FANDIÑO

Estudiante Ingeniería de Sistemas e Informática  
Universidad Industrial de Santander  
[agarcia@sis25.uis.edu.co](mailto:agarcia@sis25.uis.edu.co)

#### HENRY ARGUELLO FUENTES

MPE. Ingeniero Electricista  
Universidad Industrial De Santander  
[henryarguellofuentes@hotmail.com](mailto:henryarguellofuentes@hotmail.com)

### RESUMEN

En el presente trabajo se implementaron dos técnicas de predicción de señales de voz con el fin de tener resultados comparativos que permitan analizarlas para tener una estimación de cual tiene características más cercanas a la señal original. Se muestra paso a paso la implementación de las diferentes técnicas, el sistema desarrollado depende del locutor, se utilizan diferentes frases de un mismo individuo para comparar la efectividad de cada una de las técnicas, las muestras de la señal predicha son reconstruidas en una sola voz para poder medir características propias de este tipo de señales.

#### PALABRAS CLAVES:

LPC, RNA, predicción, reconstrucción, señales, predicción de voz

#### 1. INTRODUCCIÓN

La predicción de señales de voz, es un campo ampliamente utilizado en la digitalización, sintetización, reconocimiento, reducción de datos, envío de señales de voz, y modelado de la voz, las principales aplicaciones en la predicción mediante RNA están enfocadas en la caracterización de las señales de voz aplicadas en el reconocimiento de patrones [1][7], caracterización de ciertos sonidos, reducción de los datos en la representación de la señal, análisis de ruido en las señales de voz [4],[6] y predicción de LPC para un sistema de

cancelación del discurso [3], En cuanto a predicción mediante RNA, Se destacan trabajos [2] [5] en donde se implementan funciones de base radial, perceptrón multicapa y redes realimentadas, los cuales se realizaron utilizando metodologías de predicción de lazo abierto, de lazo cerrado e híbridas, teniendo en cuenta que estas ideas se explotan dependiendo de la aplicación en la que se implementen las redes neuronales.

Sin embargo, la investigación y los estudios de este tema se han detenido debido a que las técnicas existentes muestran resultados satisfactorios para las aplicaciones en las que ha sido utilizadas, no obstante el enfoque de estas técnicas de predicción no ha sido explotado en su totalidad, se ha pensando en abordar nuevas tecnologías y aplicaciones innovadoras.

En el presente trabajo el principal interés es predecir la voz de la manera más cercana a la original de forma que un oyente no pueda apreciar la diferencia entre la voz original y la predicha, el resultado de este estudio y

la eficiencia de la predicción se utilizará en trabajos futuros para cancelar el discurso activo en el área circundante.

## 2. CAPTURA DE LA SEÑAL

La señal de voz básicamente está constituida por ondas de presión producidas por el aparato fonador humano. Con el fin de capturar este tipo de señal analógica, se utiliza un micrófono que convierte la onda de presión sonora en una señal eléctrica.

Es necesario definir la frecuencia de muestreo y la cuantificación de los bits para tener en cuenta en el momento de manipular las señales de voz, para el presente trabajo las muestras fueron tomadas con frecuencias variables entre 8kHz y 44kHz, una cuantificación de 16 bits, este muestreo se realiza debido a que la aplicación es pensada para dispositivos móviles en donde la calidad del muestreo es baja debido a la compresión que se debe realizar a los datos.

Estas muestras se recopilaron en una base de datos, cada señal tiene una duración aproximada de 3 segundos, corresponden a frases comúnmente utilizadas en conversaciones telefónicas, las muestras estudiadas pertenecen a individuos diferentes, al realizar el entrenamiento se debe tener en cuenta, que para cada persona las características de voz varían lo que influye en una buena predicción de la señal.

### 2.1. Descripción del Ambiente:

Para realizar la captura de la señal se hizo necesario tener en cuenta el entorno en el cual se grabaron las voces. Delimitando el proyecto a señales en un ambiente controlado. La toma de las muestras se realizó un estudio de grabación de la Universidad Industrial de Santander. El nivel del ruido ambiental

corresponde al aire acondicionado del lugar de grabación

<b>Tamaño:</b>	3,95 m X 2,6 m X 2,70 m
<b>Número de Persona:</b>	1
<b>Potencia del ruido Ambiental:</b>	45 dB
<b>Características micrófono:</b>	Cobalt CO 11
<b>Tarjeta de Sonido:</b>	M – Audio Firewire 1814

Tabla 1. Características del entorno de grabación

## 3. PREDICCIÓN DE VOZ MEDIANTE COEFICIENTES DE PREDICCIÓN LINEAL

La predicción lineal es una herramienta matemática que permite modelar la voz representando la envolvente espectral digital de la señal de forma comprimida, utilizando la información de un modelo lineal. Es una de las técnicas más potentes del análisis de voz y uno de los métodos más útiles para codificar voz con buena calidad a una tasa de bits relativamente baja.

### 3.1. Procedimiento para la predicción:

4. Se determina la frecuencia de la señal, generalmente es de 8Khz y el número de coeficientes (p) a hallar por cada ventana, el tamaño de la ventana es de 120 para este estudio.
5. Cálculo de los coeficientes de predicción lineal ( $a_k$ )
  - 5.1. Se determina el valor de M (constante que indica cada cuantas muestras se calculan los coeficientes, es el corrimiento del vector  $a_k$ )
  - 5.2. Por cada segmento de la señal se calculan p valores de  $a_k$ , por el método de auto correlación regresiva para encontrar dichos coeficientes.
6. Predicción de la señal:

Para predecir n muestras se utiliza la siguiente sumatoria:

$$S[n] = \sum_{k=1}^p a_k s[n-k] \quad (1)$$

Cuyos parámetros son:  
 $s[n-1]$ , muestras anteriores del valor siguiente a calcular,  
 $P$  que representa el tamaño de la ventana, y  $a_k$  son los coeficientes de predicción.

Para predecir el primero y segundo conjunto de tamaño  $M$  de muestras se utiliza  $a_k[1]$ , mientras se llena el búfer de datos que permitan calcular  $a_k[2]$ , que permitirán calcular el tercer grupo de  $M$  muestras a partir de este momento el conjunto  $M$  se halla con  $a_{n-1}[p]$  hasta llegar al tamaño de la señal.

Implementado el algoritmo se procedió a probarlo con la variación de los parámetros,  $M$  (cuantas muestras hallar) y  $p$  (coeficientes de predicción lineal), las pruebas se realizaron para 10 señales de voz de un mismo individuo, donde se evaluó el tiempo de procesamiento requerido, el error cuadrático medio resultante para cada señal, obteniendo valores de error cercanos a cero lo que indica que LPC proporciona un buen estimativo porque la señal predicha se acerca a la original. La apreciación del error se hizo mediante la fórmula del error cuadrático medio:

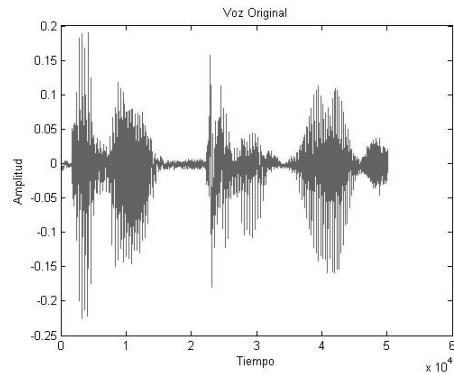
$$\sqrt{\frac{\sum_{n=1}^n (x_n - \hat{x}_n)^2}{n(n-1)}} \quad (2)$$

En donde  $x_i$  corresponde a la señal original,  $\hat{x}$  a la predicha y  $n$  es la longitud de la señal original.

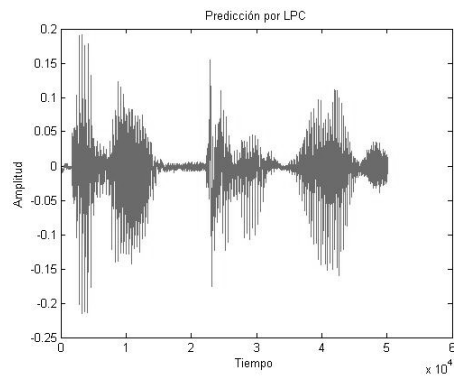
Finalmente, se encontró que la mejor arquitectura corresponde a  $M=16$  y  $P=16$ , cuyo error es igual a  $9,6 \times 10^{-3}$ ,

valor que tiende a cero lo que significa que es un buen estimativo para la predicción. El tiempo de procesamiento es igual a 54 segundos valor por debajo de la media del tiempo de procesamiento de todas las pruebas realizadas.

El resultado gráficamente con la frase: “Alo, buenos días”, se muestra en las siguientes figuras:



**Figura 1a:** Señal original correspondiente a la frase “Aló, Buenos días”



**Figura 1b:** Señal predicha con LPC correspondiente a la frase “Aló, Buenos días”

Analizando las figuras 2 y 3 es notable que la predicción por LPC muestra resultados satisfactorios, gráficamente son dos señales muy similares y al escucharlas el receptor no percibe la diferencia entre una y otra.

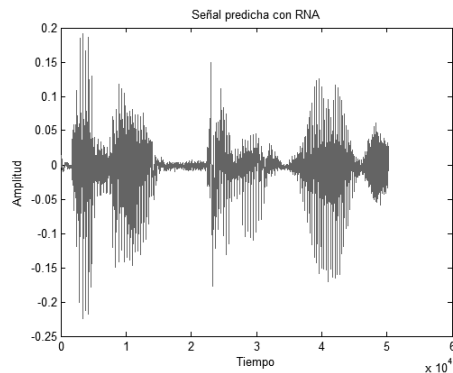
#### 4. PREDICCIÓN DE VOZ MEDIANTE REDES NEURONALES ARTIFICIALES

La predicción por medio de RNA consiste en entrenar una red con un conjunto de muestras de determinada señal de manera que retorne el siguiente dato.

El conjunto de datos de entrada es una matriz de  $20 \times n$  muestras donde  $n$  es la longitud de la señal, lo cual indica que por cada 20 muestras devuelve la muestra 21. Este proceso se realiza con una señal de longitud variable según la frecuencia de la señal de manera que la red tenga conjunto de entrenamiento bueno para usar en la predicción. La salida de la red es equivalente a la longitud de la señal.

Después de realizar múltiples entrenamientos, buscando los mejores pesos y la mejor configuración se obtuvo que la red con arquitectura de 2 capas ocultas y una de salida con 4, 9 y 1 neuronas respectivamente en cada capa, esta arquitectura se optimizó variando el algoritmo de entrenamiento y el número de épocas para obtener finalmente una configuración que garantizara buenas prestaciones relacionadas con el error de generalización y el tiempo de convergencia.

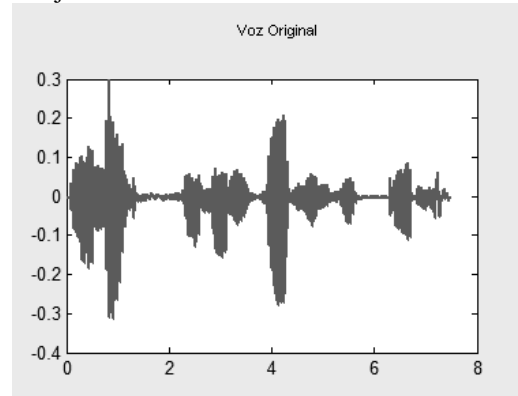
Para observar el desempeño de la red escogida, se muestran los resultados al simular con la frase 'Aló, buenos días'



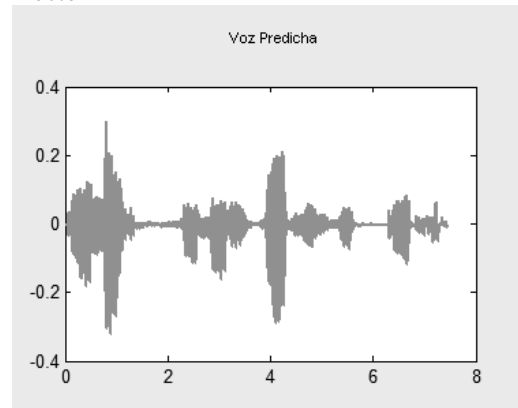
**Figura 2:** Señal predicha por RNA correspondiente a la frase "Aló, Buenos días"

## 5. COMPARACIÓN ENTRE LAS DOS TÉCNICAS

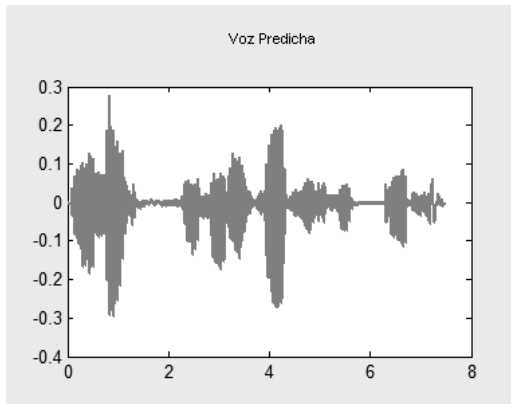
Una vez implementadas las técnicas se realizó un análisis comparativo de los resultados arrojados.



**Figura 3a** Voz original correspondiente a la frase 'Hola, buenas tardes Doctor'



**Figura 3b.** Voz predicha mediante LPC correspondiente a la frase 'Hola, buenas tardes Doctor'



**Figura 3c.** Voz predicha mediante RNA correspondiente a la frase ‘Hola, buenas tardes Doctor’

Las dos técnicas de predicción arrojaron resultados satisfactorios para los objetivos propuesto, el principal de ellos: predecir la señal de voz, la percepción acústica que se logra no permite la distinción entre las frases, gráficamente las 3 voces indiscutiblemente se ven similares excepto en algunos pedazos que son mínimos al compararlos con toda la señal en general, numéricamente los resultados son:

Calculos Generales			
	Voz Original	Voz LPC	Voz RNA
Error cuadrático medio	0.0	0.0001037	0.00052904
Frecuencia fundamental	182.4089	173.201	119.9081
Potencia	0.002268	0.0021568	0.0014913
Coefficiente de Correlacion	1	0.97358	0.96457

**Figura 4.** Resultados numéricos.

Para realizar las pruebas se tomaron 5 frases distintas de 5 oradores diferentes, prediciendo para cada una de ellas por los dos métodos. En las tablas 2, 3 y 4 se muestran los resultados para un individuo según las siguientes medidas:

1. El error cuadrático medio (E2M) se calcula al igual que los resultados obtenidos evaluando cada técnica como en la ecuación (1).

Tabla 2 Errores variando frases

Frase	E2M	
	Voz LPC	Voz RNA
1	0.00025565	0.0019252
2	0.00017862	0.00058433
3	0.00021554	0.00081281
4	0.00032277	0.00092284
5	0.00011521	0.00011272
prom	2.1756e-004	8.7158e-004

Dependiendo de la frase y del locutor el error varia, para este estimativo las dos técnicas el promedio de error tienen el mismo orden.

2. La frecuencia fundamental corresponde a la frecuencia más baja del espectro tal que las frecuencias dominantes pueden expresarse como múltiplos de esta, como puede observarse (en la tabla 3 y figura 4)

Tabla 3. Frecuencia Fundamental de las señales

Frase	Frecuencia Fundamental		
	Voz Org.	Voz LPC	Voz RNA
1	53.8529	41.3544	44.3975
2	115.3704	100.9514	97.4676
3	189.5602	163.2166	179.942
4	258.0925	213.1584	247.4627
5	157.8963	129.511	143.9003

Los resultados son similares entre sí pero numéricamente hablando la frecuencia fundamental de la voz predicha mediante RNA se acerca más a la de la voz original que la de predicción por LPC, la diferencia entre RNA y voz original es menor que entre LPC y la voz original, entre menor sea la diferencia más cercanas están una voz de la otra.

3. La potencia de las señales de voz representada por la ecuación (3), es una

medida que también tiene proximidad entre las distintas señales. Al observar los diferentes valores (tabla 4) se puede concluir que, al igual que con el caso de la frecuencia fundamental, la diferencia con RNA es menor que con LPC.

$$\frac{1}{N} \sum_{i=1}^N |s_i|^2 \quad (3)$$

N: Longitud de la señal

Tabla 4. Potencia de las señales

Frase	Potencia		
	Voz Org.	Voz LPC	Voz RNA
1	0.014224	0.011287	0.011789
2	0.036649	0.033383	0.03116
3	0.031741	0.027948	0.030232
4	0.035851	0.030141	0.03447
5	0.037143	0.031374	0.034011

Un análisis gráfico importante fue el espectrograma como se muestra en la figura 5

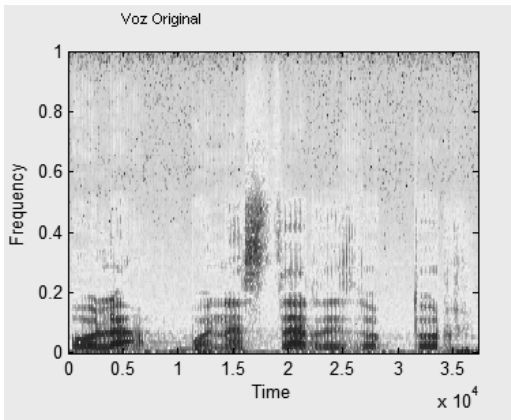


Figura 5a. Espectrograma de la señal original

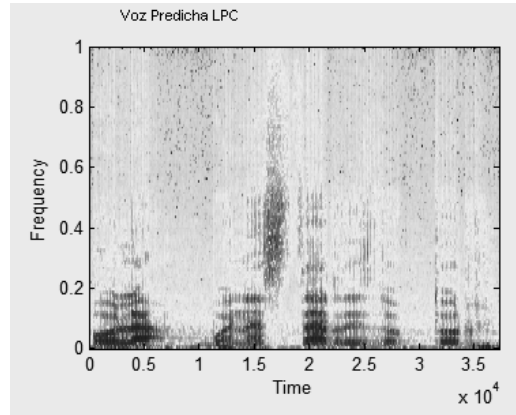


Figura 5b. Espectrograma de la señal predicha por LPC

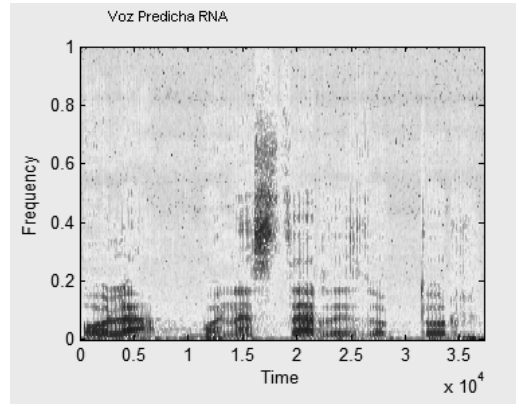
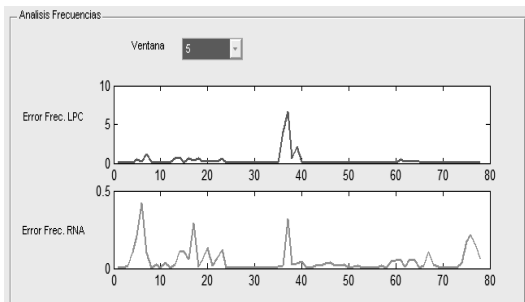


Figura 5c. Espectrograma de la señal predicha mediante RNA.

El espectrograma permite analizar la sonoridad, duración y estructura de los formantes (timbre), intensidad, pausas y ritmo, comparando los resultados gráficos se puede observar que los espectrogramas de las señales predichas con respecto a la señal original muestran que la voz predicha por RNA tiene mejores resultados que la voz predicha mediante LPC. Las frecuencias altas y bajas se encuentran aproximadamente iguales.

Para corroborar los resultados visuales y obtener un error numérico de valores de frecuencia se promedió la medida de las frecuencias mostrada en el espectrograma con ventanas de ancho variables de 1 a 10 columnas de datos.

Se demuestra que las señales tienen valores de frecuencia aproximados, en los mismos rangos de valores. Sin embargo, se presentan picos en los cuales la diferencia es considerable. En la figura 6 se muestra el comportamiento para la frase “Hola, buenas tardes doctor”, con una ventana de ancho igual a 5 donde se puede observar un pico en la grafica de error. Además, el valor de este pico para la señal predicha con LPC es mayor, la señal predicha con RNA tiene mayor aproximación en frecuencia con la original, como era de esperarse después de observar los cálculos de la frecuencia fundamental.



**Figura 6.** Análisis frecuencias por ventana

### CONCLUSIONES:

- ✓ El método de predicción lineal es comúnmente utilizado por su simplicidad, se sacrifica calidad logrando una predicción aceptable por otra parte, el método de predicción no lineal corresponde a sistemas más complejos y exige por tanto mayor esfuerzo computacional, a cambio se pueden variar mayor número de parámetros.
- ✓ La técnica de Redes Neuronales Artificiales demanda mayores recursos computacionales debido a los procesos que tiene que realizar, sin embargo, al evaluar los

resultados devueltos por los dos métodos de manera detallada se observa que en la mayoría de los casos ambos errores tienen el mismo orden e incluso en ocasiones la señal predicha con RNA es más aproximada. Por otra parte, si se hace una representación de frecuencia de la señales se puede observar que las muestras de RNA se acercan mas a la original que las de LPC. Lo cual se confirma al calcular el valor de la frecuencia fundamental, parámetro determinante en toda señal, cuyo valor está más acertado con RNA.

- ✓ El sistema funciona perfectamente cuando se evalúa con el mismo locutor, sin embargo, la RNA supera los resultados esperados, se probó con otra persona diferente a la del conjunto de entrenamiento y la generalización de la red neuronal encontrada llevo a resultados satisfactorios.
- ✓ Las pruebas realizadas para los algoritmos de predicción requieren tener en cuenta la población, las frases a utilizar el ruido ambiental y la calidad de las señales de voz. Al realizar pruebas cortas con frecuencias altas y bajas, se logra percibir la diferencia, concluyendo dos cosas: primero, las frecuencias de muestreo bajas de 4KHz predicen bien pero con cierta cantidad de ruido además el tiempo de procesamiento es bajo, para frecuencias altas de 20KHz o 44KHz, se logra más nitidez en las señales, pero el tiempo de respuesta es más largo y demorado.

### BIBLIOGRAFIA

- [1] GIOWACZ, Adam. y GOWACZ, Witold. Sound recognition of DC machine with application of LPC and metrics. Krakow. En: Human System Interactions, 2008 Conference, IEEE. (may 2008); p. 322 – 327
- [2] HACIOGLU, K. y TURKYILMAZ, R. On the use of radial basis function networks for nonlinear speech processing. Tel-Aviv, Israel. En: Electrotechnical Conference, IEEE (may. 1998); p. 29 - 33
- [3] KONDO, Kazuhiro y NAKAGAWA, Kiyoshi. On Long Term Prediction for Active Cellular Speech Emission Control. En: Proceedings of the Internoise. Korea. (aug. 2003); p. 26-15
- [4] TRABELSI, Boyer. SAVARIA, Boukadoum, Improving LPC analysis of speech in additive noise.

Montreal. En: Circuits and Systems for Communications, IEEE (agosto 2007); p. 93 – 96

[5] WANG, An-Hong. SUN Zhi-Yi. LI, Zhi-Hong, et al. A dynamic wavelet neural network for prediction of speech signal. Taiyuan. En: Machine Learning and Cybernetics, IEEE (agosto 2004); p. 3524- 3528

[6] ZHANG, Rubo. TIAN, Wu. XUEYAO, Li et al. Speech stream detection in strong noise based on linear prediction. Singapore. En: Industrial Electronics and Applications, IEEE. (mayo 2006); p. 1 - 4

[7] ZHANG, Xueying. y GUO, Yueling. HOU, Xuemei. A speech recognition method of isolated words based on modified LPC cepstrum. Fremont. En: Granular Computing,