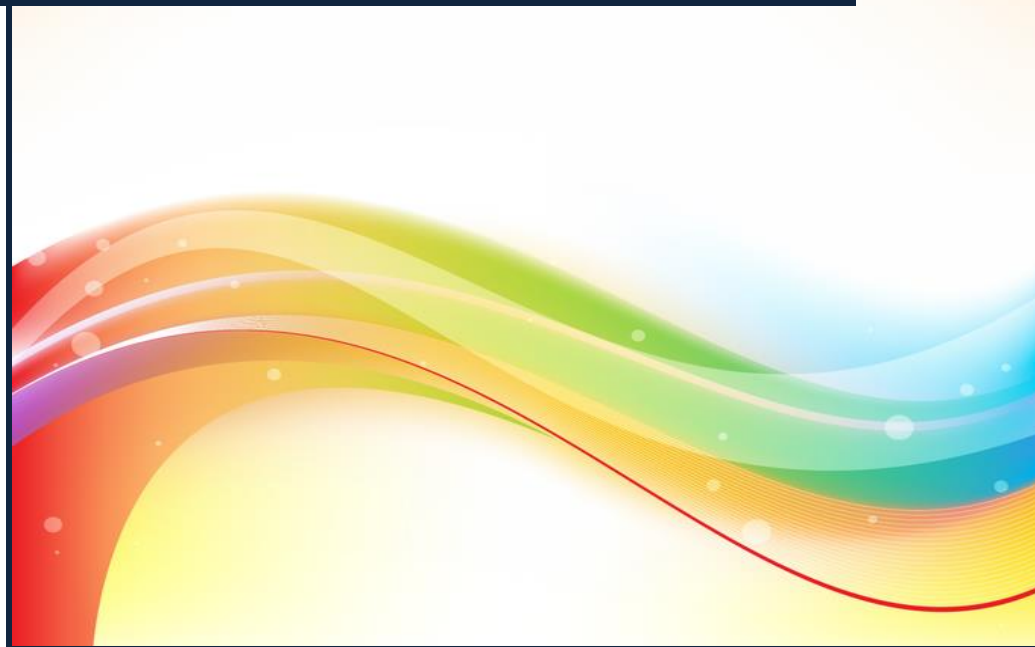


2020

**DISEÑO E IMPLEMENTACIÓN DE UN MODELO
FUNCIONAL DE GENERACIÓN DE VOZ POR MEDIO DE LA
CAPTACIÓN DE VIBRACIÓN DE CUERDAS VOCALES DE
PERSONAS CON DISCAPACIDAD DE COMUNICACIÓN
ORAL**



Juan M. Silva Rojas

Julián C. Ortiz

Jiménez

Universidad Santo Tomás
Bogotá D.C.

**DISEÑO E IMPLEMENTACIÓN DE UN MODELO FUNCIONAL DE
GENERACIÓN DE VOZ POR MEDIO DE LA CAPTACIÓN DE VIBRACIÓN DE
CUERDAS VOCALES DE PERSONAS CON DISCAPACIDAD DE
COMUNICACIÓN ORAL**



UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

JUAN MANUEL SILVA ROJAS
Código 2176830
JULIÁN CAMILO ORTIZ JIMÉNEZ
Código 2174266

UNIVERSIDAD SANTO TOMÁS
DIVISIÓN DE INGENIERIAS
FACULTAD DE INGENIERÍA ELECTRÓNICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
BOGOTÁ D.C
2020

**DISEÑO E IMPLEMENTACIÓN DE UN MODELO FUNCIONAL DE
GENERACIÓN DE VOZ POR MEDIO DE LA CAPTACIÓN DE VIBRACIÓN DE
CUERDAS VOCALES DE PERSONAS CON DISCAPACIDAD DE
COMUNICACIÓN ORAL**



UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

**TRABAJO DE GRADO para optar al título de
INGENIERO ELECTRÓNICO**

**JUAN MANUEL SILVA ROJAS
JULIÁN CAMILO ORTIZ JIMÉNEZ**

Director
OSCAR MAURICIO GELVEZ LIZARAZO
Maestro en Ciencias en la Especialidad de Electrónica

Codirector
CARLOS JAVIER MOJICA CASALLAS
Magister en Ingeniería Electrónica

UNIVERSIDAD SANTO TOMÁS
DIVISIÓN DE INGENIERIAS
FACULTAD DE INGENIERÍA ELECTRÓNICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
BOGOTÁ D.C
2020

Nota de aceptación:

**Aprobado por el Comité de Grado de la
Facultad de Ingeniería Electrónica en
cumplimiento de los requisitos exigidos
por la Universidad Santo Tomás para
optar al título de Ingeniero Electrónico**

Jurado

Jurado

Bogotá D.C. 15 de abril de 2020

DEDICATORIA

Esta tesis va dedicada a mis queridos padres que con su perseverancia y esfuerzos me permitieron alcanzar este momento, además de forjar los cimientos de mi desarrollo como persona y profesional con su paciencia afecto y sabiduría, a mi hermano mayor por ser un compañero de aventuras y un modelo a seguir con su incasable apetito de conocimiento y su responsabilidad, a mi hermana menor por estar siempre apoyándome y ayudándome a ser mejor persona.

Juan Manuel Silva.

DEDICATORIA

Dedico esta tesis a mis padres Octavio Ortiz y Esther Jiménez que fueron los pilares para cumplir mis objetivos, demostrando que siempre puedo entregar lo mejor de mí, a mi hermano Rubén que con su paciencia me demostró que puedo contar con su ayuda en cualquier momento y a mi hermano Johann por brindarme la serenidad que se debe tener al solucionar un problema para así superar todos los retos que se me presentaron a lo largo de la carrera.

Julián Ortiz.

AGRADECIMIENTOS

Quiero agradecer a todos y cada uno de los profesores que han tenido la paciencia y dedicación de educarme no solo en los temas académicos, a mi director de tesis por las horas de dedicación al proyecto, a mi compañero Julián Ortiz por su compañía, ayuda, esfuerzos y dedicación en este dispendioso y laborioso proceso.

Al neurocirujano Libardo E. Pulido, al otorrinolaringólogo Diego Chow Maya por asesorarnos en los temas médicos del proyecto, así como a la fonoaudióloga Sandra Barbosa Rincón por asesorarnos en los temas lingüísticos del proyecto, al profesor Oscar Gelvez por guiarnos en este arduo proceso

Juan Manuel Silva.

AGRADECIMIENTOS

A mis padres y hermanos ya que, gracias a su paciencia, su amor, y su comprensión me dieron la fuerza necesaria para superar los retos que constantemente se me presentaron a lo largo de la carrera, brindándome enseñanzas claves como la calma y el coraje.

A mis compañeros de universidad y compañero de tesis Juan Manuel Silva porque gracias a su amistad y trabajo en equipo fue posible afrontar los grandes retos presentes en la Universidad y reto a la vez.

Julián Ortiz Jiménez.

CONTENIDO

	Pág.
1 RESUMEN.....	14
2 PLANTEAMIENTO DEL PROBLEMA.....	15
3 ANTECEDENTES.....	16
4 JUSTIFICACIÓN.....	18
5 IMPACTO SOCIAL.....	19
6 OBJETIVOS.....	20
6.1 OBJETIVO GENERAL.....	20
6.2 OBJETIVOS ESPECÍFICOS.....	20
7 MARCO TEÓRICO.....	21
7.1 EL APARATO FONADOR.....	21
7.2 FISIOLOGÍA DE LA VOZ HUMANA.....	22
7.3 FORMANTES.....	22
7.4 LÓBULO PARIETAL IZQUIERDO.....	24
7.5 RADIO NECROSIS.....	25
7.6 PROCESAMIENTO DIGITAL DE SEÑALES DE AUDIO.....	25
7.7 TRANSFORMADA DE FOURIER.....	25
7.8 VENTANA DE HAMMING.....	26
7.9 ESPECTROGRAMA.....	27
7.10 REDES NEURONALES.....	27
8 DISEÑO METODOLÓGICO.....	29

8.1	<i>DIAGNOSTICO DEL PACIENTE</i>	29
8.2	<i>METODO PLANTEADO</i>	29
8.3	<i>SELECCIÓN DE TARJETA DE PROCESAMIENTO</i>	30
8.4	<i>SELECCIÓN DE SENSOR</i>	31
8.5	<i>DISEÑO DE LA GARGANTILLA</i>	33
8.6	<i>DISEÑO DE PROTOTIPO</i>	34
8.7	<i>ANÁLISIS Y PROCESAMIENTO</i>	42
8.7.1	<i>Análisis de muestreo.</i>	42
8.7.2	<i>Procesamiento de audio.</i>	42
8.7.3	<i>Método de solución del proyecto.</i>	50
8.7.4	<i>Diseño de la red neuronal de entrenamiento.</i>	51
9	<i>ANÁLISIS DE RESULTADOS</i>	57
10	<i>CONCLUSIONES</i>	64

ÍNDICE DE FIGURAS

	Pág.
Figura 1. Diagrama del aparato fonador.	21
Figura 2. Mediciones de los formantes de las vocales del español.	23
Figura 3. Lóbulo parietal izquierdo.	24
Figura 4. Ventana de Hamming.	26
Figura 5. Espectrograma.	27
Figura 6. Sensor Piezoeléctrico	32
Figura 7. Micrófono Electret	Error! Bookmark not defined.
Figura 8. Diseño de la gargantilla	34
Figura 9. Diagrama prototipo de captación de audio.....	34
Figura 10. Esquemático PCB	35
Figura 11. Resultado simulación filtro pasa bajos	36
Figura 12. Medición de la señal en -3db	37
Figura 13. Esquema de filtro implementado.	37
Figura 14. Tarjeta de audio USB.	38
Figura 15. Tarjeta Raspberry Pi 3B+.	39
Figura 16. Amplificado de Audio TEA2025B	40
Figura 17. Parlante.	40
Figura 18. Esquemático del prototipo	41
Figura 19. Señal original.	43
Figura 20. Señal reajustada con valores positivos.	44

Figura 21. Delimitando señales de voz dentro de los audios.	45
Figura 22. Recorte de la señal producida por una persona sana.	46
Figura 23. Señal producida por el paciente.	47
Figura 24. Transformada de Fourier de la señal de persona sana.....	48
Figura 25. Transformada de Fourier de la señal de voz del paciente.	48
Figura 26. Transformada de Fourier del paciente en 3d.	49
Figura 27. Ventana de Hamming para señal de una persona sana.	58
Figura 28. Ventana de Hamming para señal producida por el paciente	59
Figura 29. Espectrograma de la señal producida por una persona sana.	61
Figura 30. Espectrograma de la señal producida por el paciente.	62
Figura 31. Diagrama de formantes producidos por el paciente (vocal E).	63

ÍNDICE DE TABLAS

	Pág.
Tabla 1. Determinación anatómica de las características de la voz.	23
Tabla 2. Posición de la lengua al producir los fonemas de las vocales.	24
Tabla 3. Comparación de tarjetas de procesamiento	30
Tabla 4. Características del micrófono seleccionado	Error! Bookmark not defined.
Tabla 5. Medidas de la gargantilla	33
Tabla 6. Componentes utilizados en el prototipo.	35
Tabla 7. Cantidad de muestras por vocal.	42
Tabla 8. Cantidad de muestras por vocal.	50
Tabla 9. Pesos característicos del target para la red.	51
Tabla 10. Pruebas de la red neuronal con las muestras del paciente.	57

ÍNDICE DE CÓDIGOS

	Pág.
Código 1. Implementación en código de la ecuación 2	43
Código 2. Hallando las señales de voz dentro de la muestra de audio.	45
Código 3. Librerías de la red neuronal.	51
Código 4. Creación del modelo de red neuronal.	52
Código 5. Optimizador de la red neuronal.....	53
Código 6. Método de entrenamiento	53
Código 7. Modelo de almacenamiento de datos.	54
Código 8. Carga del modelo de predicción y los pesos	54
Código 9. Carga de la señal de entrada en el modelo de predicción.....	55
Código 10. Resultado de la clasificación de la red.....	55
Código 11. Implementación de texto a voz de la librería pytsx3	55
Código 12. Estimación de coeficientes LPC.	60

1 RESUMEN

Este documento muestra el estudio, simulación e implementación de un modelo de generación de voz por medio de la captación de vibración de cuerdas vocales de personas con discapacidad de comunicación oral.

Este proyecto tuvo el objetivo de entender y dar solución a un problema en particular de una persona con dificultades del habla, para esto fue necesario investigar sobre el diagnóstico del paciente y con base a esto se implementó un dispositivo el cual consta de micrófonos situados al costado de la garganta para captar las vibraciones de las cuerdas bucales y posteriormente realizar el debido procesamiento mediante herramientas como transformada de Fourier y métodos de redes neuronales y filtros adaptativos.

2 PLANTEAMIENTO DEL PROBLEMA

La comunicación es una de las necesidades básicas del ser humano, la cual ha permitido el desarrollo del mismo en sociedad, siendo el habla el mecanismo más importante, sin embargo; existe un grupo de personas que no cuentan con esta forma de interacción por diferentes causas, como pueden ser: deficiencias en los mecanismos de comunicación desde el nacimiento, como la parálisis cerebral que es un trastorno progresivo no permanente que afecta el tono muscular, el movimiento y las habilidades motoras; un accidente a lo largo de su vida, capaz de ocasionar la pérdida del habla, como el traumatismo craneoencefálico que de ser grave, puede causar el deterioro de la función del cerebro, la pérdida de la capacidad para hablar o escuchar, dependiendo de la gravedad del golpe; un mal procedimiento quirúrgico que afecte la boca o el hemisferio cerebral encargado de la capacidad del habla; también se puede considerar una infección o un virus que agravan este problema como lo puede ser la Meningitis bacteriana la cual es una inflamación en las capas del tejido que cubren el encéfalo lo que a su vez provoca una rigidez en el cuello dificultando su movilidad, además de producir sordera en algunos casos.

El presente proyecto se enfoca en un paciente afectado por **radio necrosis cerebral**, enfermedad que deteriora determinados lóbulos del cerebro o la comunicación entre este y otras áreas del cuerpo, lo que imposibilita o dificulta algunas acciones. En algunos casos es producida por la radioterapia o quimioterapia utilizada en el tratamiento posterior a la extirpación de un tumor. En este caso la **radio necrosis cerebral** específicamente **frontoparietal izquierda** que afecta la zona encargada de la comunicación y los movimientos musculares generados en la cavidad bucal lo que dificulta procesos como la deglución que es el proceso por el cual una persona transporta alimentos o saliva desde la cavidad bucal hasta el estómago y la fonación, que hace referencia a la voz hablada.

El paciente no presenta afectación en las cuerdas bucales, las cuales funcionan correctamente y producen sonido, pero al llegar a la cavidad bucal esta no resuena ni tiene acentuación lo que le impide la comunicación normal con otras personas.

Debido a lo anterior, se plantea el diseño de un modelo funcional que procese las señales obtenidas de sensores situados al costado del cartilago cricoides, generadas por las cuerdas bucales de una persona.

3 ANTECEDENTES

K. Ishizaka y J.L. Flagan publicaron un artículo en el año 1972 en Nokia Bell [1] en el cual se realizó un modelo de circuito equivalente de la glotis, permitiendo modelar matemáticamente su comportamiento, facilitando el entendimiento sobre la conducta de la señal producida al pasar el aire por esta, permitiendo la vibración de las cuerdas generando el sonido, también permite entender qué función cumple esta cavidad en temas de resonancia de la señal de la voz. A partir de lo anterior para conocer la implementación de modelos matemáticos a una señal generada por una persona con dificultades de habla, se estudia el artículo del año 2009 de Lodz Polonia escrito por Damian krzesimowski y Zigmunt Ciota [2], en el cuál una vez transcritos los algoritmos matemáticos en programas especializados que permiten observar las formas de procesamiento que mejoren o aproximen la señal del paciente a la esperada, facilita conocer las posibles metodologías para el tratamiento de las señales obtenidas de los sensores, pero teniendo en cuenta que las personas que produjeron las señales en este proyecto poseían el aparato fonador en óptimas condiciones.

Por otra parte en el artículo del 2010 publicado en IEEE Transactions on Microwave Theory and Techniques escrito por Chien-San Lin, Sheng-Fuh Chang, y Chia-Chan Chang [3] se aborda un proyecto, donde extraen información de voz de la señal de vibración que producen las cuerdas bucales, mediante un sistema de radar de voz, el cual permite a partir de un circuito electrónico detectar el movimiento glótico, además de realizar mediciones que mejoran la sensibilidad y calidad en la detección de la señal, el cual podría servir de referencia para realizar mediciones que mejoren la señal capturada por los sensores.

Complementando los modelos matemáticos encontrados en el artículo del año 2009, V. S. Balaji, N. R. Raajan y Har Narayan Upadhayay publicaron en 2012 en Sivakasi India [4], un modelo matemático general que permite conocer la forma y características del aparato bucal, indicando posibles patrones presentes de forma constante en la generación de determinadas letras, que de estar presentes en el paciente permiten conocer mejor el comportamiento de la señal sin realizar un procedimiento invasivo que incrementa las complicaciones al paciente con el objetivo de llegar posiblemente a una misma conclusión.

Adicional a lo desarrollado por los diferentes autores mencionados anteriormente, es de gran importancia conocer la forma en que se produce un sonido: la frecuencia del pulso global (GRP) que determina el tono vocal es un trabajo realizado por Martin Andermann, Roy D. Patterson, Carolin Vogt, Lisa Winterstetter y André Rupp en la publicación del 2017 en NeuroImagen [5], en el cual se explican las propiedades de la voz, técnicas inmunes a la variación de forma de onda asociada, y cuál fue el procedimiento a seguir para concluir dicho estudio. El cuerpo humano emite el sonido al vibrar las cuerdas bucales por un impulso del aire, como se puede

apreciar en el artículo [6], además es posible conocer su comportamiento debido al artículo del año 2017 de Turín Italia escrito por F. Casassa, A. Carullo y A. Vallan [7] donde se diseñó un sistema que simula el aparato fonador mediante un dispositivo electrónico, describiendo el proceso de vibración de las cuerdas vocales y como se da la señal por medio de la vibración, que normalmente no se puede apreciar. Lo anterior ofrece una ventaja para el desarrollo del proyecto al conocer la señal de salida en relación a la cantidad y velocidad de vibraciones y de esta forma describir el comportamiento de las cuerdas bucales.

Finalmente, en el artículo del año 2018 publicado en Biomedical Signal Processing and control escrito por Susana Vaz-Freitas, Pedro Melo Pestana, Víctor Almeida y Aníbal Ferreira [8], se definen los parámetros que caracterizan los atributos físicos de la voz en el tiempo, ya que una señal presenta varias características acústicas, dentro de las cuales las más importantes son: periodicidad, amplitud, duración y composición espectral. Para lograr dicha conclusión se realizaron pruebas bucales a 90 personas las cuales 21 se consideraron normales (23,30%) y 69 mostraron un grado de trastorno, 20 un trastorno leve en la voz (22,22%), 47 moderado (52,22%) y 2 lesiones severas (2,22%). Los autores usaron el mismo protocolo para registrar las señales de la voz: frecuencia de muestreo de 44100 Hz, 16 bits por muestra, unidireccional.

A partir de la información recopilada y presentada en este documento, para desarrollar un diseño e implementación de un modelo funcional de generación de voz por medio de la captación de vibración de cuerdas vocales de personas con discapacidad de comunicación oral, se plantea una solución considerando la importancia de la frecuencia y la resolución en el procesamiento de la señal con el fin de obtener un buen resultado, además de encontrar parámetros acústicos como el Jitter para la comparación entre las señales del paciente y las de una persona sana, permitiendo dar una solución práctica a personas con una discapacidad específica.

4 JUSTIFICACIÓN

El cáncer es la segunda causa de muerte más frecuente a nivel mundial después de los ataques cardíacos [9] se diagnostican alrededor de 23,880 tumores malignos (13,720 en hombres y 10,160 en mujeres), de estos aproximadamente 16,830 morirán (9,490 hombres y 7,340 mujeres). Los tratamientos que son utilizados para combatir el cáncer son: la Quimioterapia y Radioterapia, en esta última se utilizan rayos x, que afectan todas las células lo que puede ocasionar daños a zonas no comprometidas, generando enfermedades como **radio necrosis cerebral**, enfermedad que disminuye la actividad cerebral en los hemisferios afectados, que al ser crónica y no poseer tratamiento médico que mejore la condición del paciente, surge la necesidad de crear dispositivos que suplan dichas deficiencias, por ejemplo el caso de Julio César Convers Sánchez quien posee esta enfermedad específicamente del tipo **frontoparietal izquierda** (según historia clínica), afectando su capacidad de habla. Actualmente el acceso a pacientes con patologías clínica específicas cuenta con dificultades logísticas y legales debido a factores como la confidencialidad médico-paciente, sin embargo, en el presente proyecto se cuenta con la facilidad de trabajar con Julio César Convers Sánchez a partir de su deseo de entablar nuevamente conversación con otras personas de forma oral.

Debido a esto, se busca diseñar un sistema que reduzca las dificultades de habla del paciente, desde el punto de vista de la ingeniería electrónica, con las tecnologías disponibles en el mercado, además de los programas licenciados presentes en la Universidad Santo Tomás, cuya arquitectura tengan la capacidad de procesamiento necesaria para el proyecto.

El sistema se diseña para ser lo más cómodo posible para el paciente, debido a que no se implementan sensores de inserción puesto que el paciente presenta problemas en el proceso de deglución y estos dificultan dicho proceso, además de ocasionar otras afecciones, por lo que se optaría por sensores de superficie que son pequeños, portables y pueden ser retirados sin mayor dificultad en cualquier momento, razón por la cual no generan presión en la garganta del usuario ni molestias adicionales, estas características permiten que el sistema se pueda adaptar a personas con enfermedades de diagnósticos similares.

5 IMPACTO SOCIAL

La comunicación oral es una de las formas de expresión más importantes en el ser humano, ya que permite compartir experiencias e ideas con diferentes personas. La limitación en la comunicación es un problema social muy importante, ya que genera exclusión de las personas que la padecen. Con este proyecto se busca generar soluciones que permitan a pacientes específicos poder comunicarse de forma independiente nuevamente. El proyecto busca brindar una solución a un paciente que no puede pronunciar palabras ni vocales, a través de elementos electrónicos que pueden ser reemplazados fácilmente y que no limitan completamente el funcionamiento del prototipo por su precio. Se busca que estas iniciativas puedan ser implementadas en una gran variedad de personas que presenten diferentes problemas fonoaudiológicos, ya que esto favorecerá a que los pacientes mejoren su calidad de vida y logren comunicarse nuevamente.

A la hora de realizar las pruebas con el prototipo se observan las dificultades, producidas por el diagnóstico del paciente a la hora de realizar procesos cotidianos como la deglución de alimentos y el habla, conociendo nuevos síntomas que no se conocían con anterioridad, así como la agrafia que es la pérdida parcial del lenguaje escrito, derivación del diagnóstico.

6 OBJETIVOS

6.1 OBJETIVO GENERAL

- ✓ Implementar un modelo de generación de voz con base en la adquisición de señales producidas por cuerdas vocales para facilitar la comunicación oral de personas con discapacidad de comunicación.

6.2 OBJETIVOS ESPECÍFICOS

- ✓ Diseñar e implementar un circuito electrónico para la captación de las señales producidas por las cuerdas bucales, de acuerdo con el diagnóstico específico del paciente (radio necrosis frontoparietal izquierda).
- ✓ Seleccionar el hardware y software más adecuado para procesar las señales vibratorias producidas por las cuerdas bucales.
- ✓ Procesar en el dominio del tiempo y la frecuencia, las señales obtenidas durante el proceso de detección de las vocales, mediante el software y hardware seleccionado para analizar las características de las señales.
- ✓ Diseñar e implementar un circuito electrónico para generar una señal de voz de las vocales a partir del procesamiento de la vibración de las cuerdas bucales.

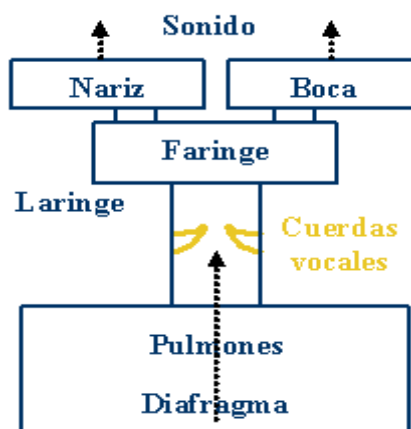
7 MARCO TEÓRICO

Para comprender con mayor detalle este proyecto, es necesario aclarar algunos conceptos claves con la finalidad de contextualizar el problema real del paciente y las diferentes herramientas utilizadas para su solución.

7.1 EL APARATO FONADOR

Según las leyes de la acústica deben existir 3 elementos indispensables para la emisión de un sonido las cuales son: el cuerpo vibrante, un medio elástico para la propagación de las vibraciones y una caja de resonancia que las amplifica, las cuales se cumplen en el aparato fonador humano (ver figura1) y se divide en tres partes fundamentales:

Figura 1. Diagrama del aparato fonador.



Fuente <http://www.ehu.es/acustica/espanol/musica/vohues/vohues1.gif>

- ✓ La cavidad infra glótica: es el lugar donde se ubican los órganos respiratorios como los pulmones.
- ✓ La cavidad laríngea o glótica: corresponde al lugar ocupado por el principal órgano fonador, la laringe y faringe.
- ✓ La cavidad supra glótica: espacio donde se localizan los órganos de la articulación y resonancia como las mejillas, la lengua entre otros [10].

7.2 FISIOLÓGÍA DE LA VOZ HUMANA

La voz se produce mediante la acción conjunta de diferentes músculos como la laringe cuya parte esencial es la glotis y los pulmones, encargados de expulsar el aire haciendo vibrar los dos pares de cuerdas vocales, cada cualidad de la voz como lo es la fonación, tono, timbre e intensidad, pueden ser modificadas con pequeñas variaciones en los esfuerzos de los diferentes músculos involucrados en la emisión de la voz, ya que la fonación puede cambiar con el cierre y la apertura constante de los pliegues vocales (cuerdas vocales), además su longitud y grosor está directamente relacionada con el tono que produce cada persona; mientras que el timbre y la intensidad dependen de la presión del aire impulsado desde los pulmones y la boca, que funciona, como caja de resonancia de acuerdo con el artículo de Dra. Begoña Torres (2007) [11].

7.3 FORMANTES

Los formantes son los puntos de amplitud máximos de resonancia de cada una de las vocales los cuales dependen, de la excitación de dos zonas llamadas F1 y F2.

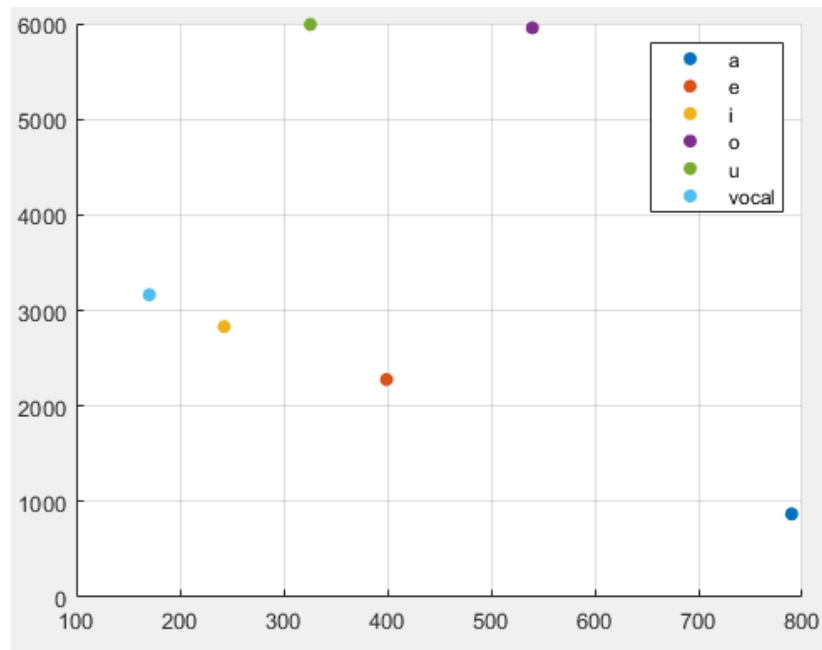
- ✓ F1 (Formante Número 1) maneja un rango de baja frecuencia entre 200 Hz a 1200 Hz aproximadamente, de igual forma se divide dicho rango en intervalos que dependen de la vocal que se escoja y de otras características propias de cada individuo.
- ✓ F2 (Formante Número 2) tiene un rango de operación superior a los 500 Hz hasta los 3500 Hz [12].

Las formantes F1 y F2 se convierten en bandas de frecuencias en donde se concentra la mayor parte de energía sonora, cuando se estimulan las cuerdas vocales. La figura 2 da una visión más detallada del concepto explicado anteriormente, acerca de los rangos en los que trabaja cada una de las vocales.

En la Tabla 1 se puede apreciar cuales componentes del aparato fonador modifican las frecuencias fundamentales y que tanto el formante 1 está relacionado a la apertura de la boca y el formante 2 a la forma de los labios.

Adicional a lo expuesto anteriormente para la producción de las vocales se debe tener una posición específica de la lengua como se observa en la Tabla 2, la cual es de vital importancia para producir los fonemas vocálicos correctamente lo cual se puede evidenciar en el paciente, ya que al no poder mover la lengua con total libertad esto afecta los sonidos característicos de la vocal.

Figura 2. Mediciones de los formantes de las vocales del español.



Fuente Propia

Tabla 1. Determinación anatómica de las características de la voz.

Formantes	Determinada por	Mayor Frecuencia	Menor frecuencia
F0	Pliegues vocales	> cantidad de ciclos/segundos	< cantidad de ciclos/segundo
F1	Apertura bucal	> apertura bucal	< apertura bucal
F2	Redondea miento/protrusión labial	+ anterior	+ posterior
		- protrusión	+ protrusión
F3	Velo de paladar	+ anterior y baja	+ posterior y alta
		Descenso velar	Elevación velar

Fuente: http://repositorio.unab.cl/xmlui/bitstream/handle/ria/1230/Cisternas_PF_Caracter%C3%ADsticas%20ac%C3%BAsticas%20de%20las%20vocales_2012.pdf?sequence=3&isAllowed=y%20

Tabla 2. Posición de la lengua al producir los fonemas de las vocales.

Vocal	Posición vertical de la lengua	Posición horizontal de la lengua
/a/	Baja	Central
/e/	Media	Anterior
/i/	Alta	Anterior
/o/	Media	Posterior
/u/	Alta	Posterior

Fuente:http://repositorio.unab.cl/xmlui/bitstream/handle/ria/1230/Cisternas_PF_Caracter%C3%ADsticas%20ac%C3%BAsticas%20de%20las%20vocales_2012.pdf?sequence=3&isAllowed=y%20

7.4 LÓBULO PARIETAL IZQUIERDO

El lóbulo parietal participa en todos los procesos sensoriales y perceptivos en el día a día, también es conocida como la corteza de asociación, ya que integra la información visual, auditiva y somatosensorial [13].

El lóbulo parietal izquierdo (ver figura 3) presenta más actividad en las personas diestras, además este hemisferio, es el encargado del procesamiento de la simbología de letras y números.

Figura 3. Lóbulo parietal izquierdo.



Fuente: <https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQRfTbJKMneXj9kPpBTRa-oR6-ydbKqzWLb0dz1PPIASz2NeQbR&usqp=CAU>

Las personas que tienen una lesión en los lóbulos parietales a menudo muestran deficiencias tales como alteraciones en la imagen corporal y las relaciones espaciales.

7.5 RADIO NECROSIS

La radio necrosis es la afectación o lesión cerebral causada por la radioterapia craneal realizada típicamente para el tratamiento de tumores o malformaciones vasculares cerebrales, esta consiste en la muerte de células específicas para controlar el crecimiento de tumores o eliminarlos [14].

Cuando la radio necrosis afecta el lóbulo parietal izquierdo puede dar lugar a lo que se llama “Síndrome de Gerstmann”, el cual incluye confusión derecha-izquierda, dificultad con la escritura (agrafia) y la dificultad con las matemáticas (acalculia), también puede producir trastorno del lenguaje (afasia) y la incapacidad para percibir los objetos (agnosia).

7.6 PROCESAMIENTO DIGITAL DE SEÑALES DE AUDIO

El procesamiento de señales es una de las tecnologías con mayor relevancia en el avance científico del último siglo y se distingue de otras áreas de las ciencias computacionales, por el tipo de datos que utiliza (señales). Las señales de audio analógicas pueden ser representadas como una secuencia de números binarios, los cuales son obtenidos al utilizar un conversor análogo-digital, de esta forma una computadora procesa dicha información, sin embargo, en el momento de la transmisión esta se ve minimizada por el proceso de conversión descrito [15].

7.7 TRANSFORMADA DE FOURIER

La transformada de Fourier es una herramienta matemática útil para convertir señales del dominio del tiempo al dominio de la frecuencia, la finalidad de esta operación matemática es permitir la simplificación de un problema en ecuaciones algebraicas [16].

Para poder aplicar la transformada de Fourier a una señal, esta debe cumplir con tres condiciones.

- ✓ Debe ser absolutamente integrable: tal como su nombre lo dice es una función cuyo valor absoluto es integrable.
- ✓ Deben tener intervalos finitos: deben existir un número finito de máximos y mínimos.

✓ Debe existir un número finito de discontinuidades.

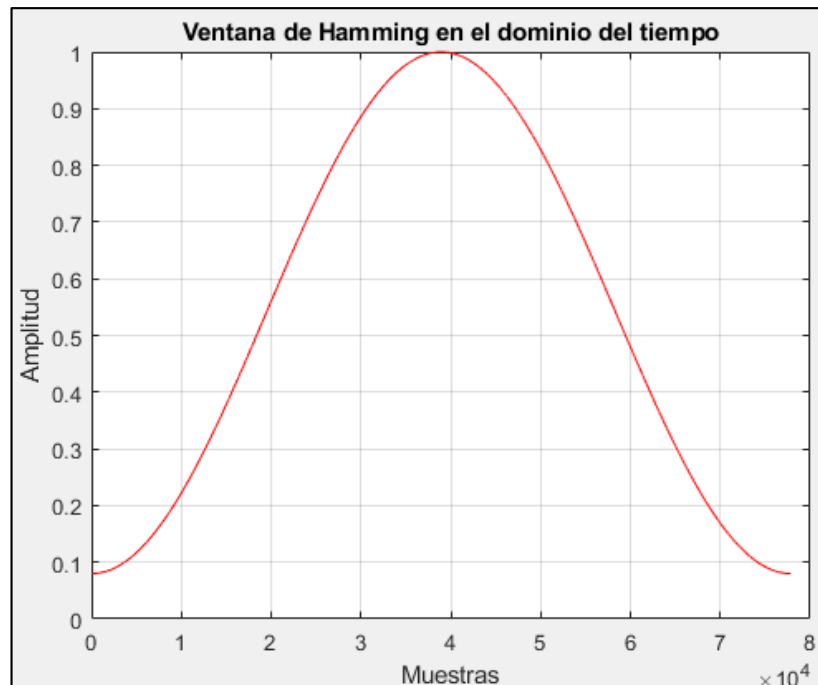
La ecuación de la transformada de Fourier se presenta a continuación en la ecuación 1

$$X(f) = \int_{-\infty}^{+\infty} X(t) * e^{-i2\pi ft} dt \quad \text{Ecuación 1}$$

7.8 VENTANA DE HAMMING

Las ventanas en el contexto de procesamiento de señales son funciones matemáticas utilizadas para evitar discontinuidades al principio y final de etapas o bloques que hacen parte de sistemas complejos; entre las múltiples ventanas que existen, la ventana de Hamming opera cancelando el lóbulo lateral más cercano, con el fin analizar una señal con una longitud limitada, en la figura 4 se observa un ejemplo de la función descrita [17].

Figura 4. Ventana de Hamming.



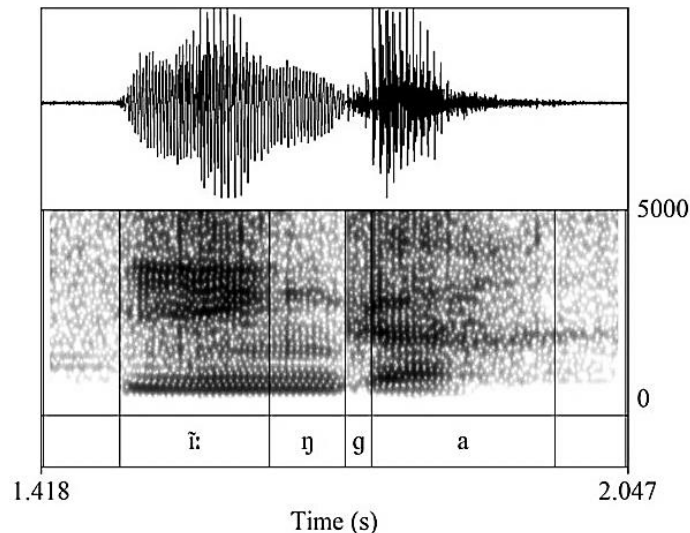
Fuente Propia

7.9 ESPECTROGRAMA

Existen rangos de frecuencias sonoras que el oído humano no es capaz de percibir, sin embargo, gracias al espectrograma es posible ver las características que se presentan en frecuencias inferiores a 20 Hz y superiores a 20 kHz, debido a que esta es una representación gráfica del espectro de frecuencias de emisión sonora [18], además este tipo de graficas es mayormente utilizado para el estudio del comportamiento de la voz humana y registros bioacústicos.

La forma en la que se obtiene un espectrograma como el observado en la figura 5 es mediante el uso de la transformada de Fourier a la señal de análisis. Dependiendo del tamaño de la ventana utilizada para el estudio de la señal, se obtiene diferentes niveles de resolución, si la ventana utilizada resulta ser muy pequeña se obtendrá en una resolución baja, ocasionando que no se pueda distinguir los armónicos de la señal.

Figura 5. Espectrograma.



Fuente

https://www.researchgate.net/profile/Braulio_Becerra_Roldan/publication/331212194/figure/fig1/AS:728127277916163@1550610479847/Figura-4-Oscilograma-y-espectrograma-de-inga-otroque-muestra-la-realizacion.jpg

7.10 REDES NEURONALES

Las redes neuronales son modelos matemáticos simplificados que simulan el procesamiento de un cerebro humano, la aplicación más común de estas redes es

el reconocimiento de patrones de cualquier tipo [19], para lo cual el procesamiento se organiza en 3 capas:

- ✓ La primera capa es la entrada, la cual recibe los datos o parámetros indicando si la neurona está activa o no.
- ✓ La segunda capa es el núcleo, en esta se representa la memoria de la red, de modo que se realizan todos los procesos necesarios para determinar la salida de la neurona, el proceso cambia según el tipo de red neuronal que se esté implementando.
- ✓ La tercera capa, por último, es la de salida que entrega la respuesta de la neurona.

Dentro de las redes neuronales se encuentra el concepto de peso sináptico que define la fuerza de una conexión sináptica entre dos neuronas, las neuronas presinápticas y las postsinápticas, estos pesos pueden tomar valores positivos, negativos o cero, con esto la red neuronal es capaz de adaptarse a cualquier entorno o realizar cualquier tarea.

Las redes neuronales cuentan con dos fases o modos de operación: La fase de aprendizaje o entrenamiento, y la fase de ejecución o prueba.

En la fase de entrenamiento se selecciona que tipo de aprendizaje va a tener la red neuronal, este puede ser: supervisado, no supervisado y por refuerzo.

- ✓ Aprendizaje supervisado: Se le presenta a la red un conjunto de patrones de entrada junto con la salida esperada. Los pesos se van modificando según el error generado por la salida real y la salida esperada.
- ✓ Aprendizaje no supervisado: Se presenta a la red un conjunto de patrones de entrada. No hay información sobre la salida esperada y los pesos serán ajustados según la correlación que tengan los datos de entrada.
- ✓ Aprendizaje por refuerzo: En esta metodología de aprendizaje se implementa las dos anteriores ya que de igual forma se le proporciona un conjunto de patrones de entrada y se le indica a la red si la salida resultante es o no la esperada, sin embargo, no se le proporciona a la red la salida esperada.

En la fase de operación se pone a prueba el aprendizaje adquirido por la neurona en el proceso anterior, presentándole datos distintos para que la red neuronal pueda encontrar los patrones presentes en la nueva información, comparándolos con los datos aprendidos anteriormente.

8 DISEÑO METODOLÓGICO

8.1 DIAGNÓSTICO DEL PACIENTE

La afección que presenta el paciente se denomina radio necrosis, la cual es la muerte de células, en este caso afecta los lóbulos frontal y parietal izquierdo, los cuales se encargan de los procesos del lenguaje como lectura, escritura y habla, además de procesos como la deglución de alimentos y el control de los músculos faciales.

En el caso del lenguaje oral el paciente presenta afasia, la cual es la pérdida de la capacidad de controlar los movimientos de la mayoría los músculos fáciles y bucales, dificultando el proceso de fonación, el cual es la capacidad de cambiar los sonidos producidos por el aparato fonador mediante el movimiento muscular de órganos como las mejillas o la lengua, gracias al cual el ser humano es capaz de diferenciar una palabra de otra.

De lo anterior cabe resaltar que el paciente tiene la capacidad de producir sonidos a partir de las cuerdas bucales, pero no posee la capacidad de realizar los fonemas característicos de cada palabra.

8.2 MÉTODO PLANTEADO

La radio necrosis frontoparietal izquierda que presenta el paciente afecta directamente el lóbulo del lenguaje y el movimiento de algunos músculos de la cara y cuello, sin afectar las cuerdas bucales del paciente lo que le permite producir sonidos, sin embargo, es la capacidad de modificar los fonemas producidos mediante la modificación de la cavidad de resonancia, así como la cantidad de aire en el proceso debido a la incapacidad de controlar los partes del cuerpo que intervienen en el proceso como la lengua, las mejillas y el movimiento de la mandíbula.

Debido a la complejidad del diagnóstico del paciente y su impacto en el proceso de fonación, se plantea el diseño e implementación de un sistema que procese las señales de audio producidas por el paciente mediante sensores no invasivos o que requieran un procedimiento quirúrgico para su implementación que puedan tener una afectación en el diagnóstico del paciente, una vez sensada la señal se envía a una tarjeta que realiza un procesamiento de la señal y posteriormente las ingresan a una red neuronal para proceder a realizar un proceso de identificación de las señales producidas, para acto seguido analizar las respuestas obtenidas.

8.3 SELECCIÓN DE TARJETA DE PROCESAMIENTO

Los diferentes sistemas embebidos que se encuentran actualmente en el mercado favorecen el desarrollo de múltiples proyectos según el propósito específico a cumplir, particularmente en el presente trabajo se consideran 12 elementos para la selección de una tarjeta que permita el procesamiento de señales de voz humana, como se aprecia en la tabla 3.

Tabla 3. Comparación de tarjetas de procesamiento

Características Tarjetas	Omega2 Onion	Raspberry Pi3	TMS320C5535
CPU	580MHz MIPS CPU	BBCM2837 ARMv8 4 núcleos a 1.2GHz	C5535 de 16 bits
Tamaño	34*20*2.8mm	82*56*19.5mm	12*12*1.4mm
Memoria RAM	128MB	1GB	256KB SDRAM
Sistema Operativo	Linux	Raspbian basado en Linux	DSP/BIOS, VLX
Entrada -Salida de audio estéreo	NO	SI	SI
MicroSD	SI	SI	NO
Puertos USB	SI	SI	SI
HDMI	NO	SI	SI
I2C	SI	SI	SI
SPI	SI	SI	SI
I2S	SI	SI	SI
UART	SI	SI	SI
Costo	\$288.890	\$389.189	\$929.000

Fuente Propia

La tarjeta que se selecciono fue la Raspberry Pi3, debido a que cuenta con una mejor CPU y una mayor capacidad de memoria RAM que las tarjetas TMS320C5535 y Omega2 Onion como se observa en la tabla 3, lo que permite realizar actividades

más complejas como lo son el cálculo de transformadas de Fourier, cálculos de coeficiente de predicción lineal (LPC), entre otros cálculos usados en el presente proyecto, adicionalmente se observa que tanto la Raspberry pi3 como la Omega2 Onion permiten trabajar sobre un sistema operativo basado en Linux, gracias a que cuenta como memoria con una tarjeta MicroSD lo que permite reemplazarla por una de mayor capacidad de acuerdo a las necesidades del diseño planteado, por esta razón tienen una ventaja frente a la tarjeta TMS320C5535 debido a que maneja un sistema operativo (DSP/BIOS, VLX) que solo permite la programación del dispositivo en lenguaje ensamblador o lenguaje C, siendo ensamblador un lenguaje de bajo nivel y C un lenguaje de medio nivel, lo que dificulta la codificación de las redes neuronales a diferencia del lenguaje Python que maneja la Raspberry Pi3 y la Omega2 Onion, ya que además de ser código abierto cuenta con librerías que contienen modelos predefinidos de redes neuronales.

Cabe resaltar que las tarjetas Raspberry Pi3 y la TMS320C5535 presentan puerto HDMI que permite la visualización de la ejecución del código en tiempo real mediante una pantalla, generando una ventaja a la hora de realizar cambios en la estructura del código.

Una de las mayores desventajas de la Omega2 Onion es la falta del puerto de entrada- salida de audio estéreo que es de suma importancia en este proyecto, puerto que si se encuentra en las tarjetas Raspberry Pi3 y TMS320C5535.

Una de las características tomadas en cuenta a la hora de seleccionar la tarjeta de procesamiento fue la relación costo beneficio, ya que según lo observado en la tabla 3, y el análisis presentado anteriormente se observa que la Raspberry Pi3 ofrece el mejor precio en relación a las características que ofrece, con la desventaja que solo cuenta con un conector estéreo de salida, a diferencia de la tarjeta Omega2 Onion, la cual es más barata pero posee una menor capacidad de CPU y memoria RAM, además de no contar con puerto de entrada-salida de audio que es de gran importancia para este proyecto, en cuanto a la tarjeta TMS320C5535 posee el mayor precio de las tres tarjetas, pero no tiene la capacidad de ampliar la memoria, y el uso de lenguajes de alto nivel.

8.4 SELECCIÓN DE SENSOR

Se planteó en un principio, el uso de un sensor piezoeléctrico de vibraciones LDT0-28K que se observa en la figura 6, que es un piezoeléctrico compuesto por un polímero laminado de plata que crea una señal a medida que el piezoeléctrico se deforma, permitiendo calcular la frecuencia producida por la deformación de los músculos al realizar el proceso de fonación, debido a que la frecuencia de la voz se encuentra entre los 80 Hz- 8.000Hz [20] cuya frecuencia está dentro del rango del sensor, pero fue descartado debido a que las cuerdas bucales funcionan con

normalidad, produciendo señales audibles lo que permite captarlas y grabarlas con micrófonos

Figura 6. Sensor Piezoeléctrico



Fuente

<https://www.electronicaembajadores.com/es/Productos/Detalle/SSPZ001/sensores/sensores-piezoelectricos/sparkfun-sen-09196-sensor-piezoelctrico-vibraciones-golpes-flexibilidad-tacto-ldt0-028k>

Para el diseño del prototipo es necesario implementar un micrófono que sea pequeño, económico y asequible de forma que pueda ser sustituido en caso de daño, por lo que se proceden a analizar 3 tipos de micrófono que cumplen con las características anteriormente mencionadas los cuales se presentan en la tabla 4.

Tabla 4. Características de micrófonos

Características Micrófonos	Electret	Micrófono de solapa	ATM350
Tipo	Condensador	Condensador	Condensador
Tipo de salida	Analógico	Analógico	Analógico
Dirección	Omnidireccional	Omnidireccional	Omnidireccional
Sensibilidad	-44dB +/- 2dB	-45dB +/- 3dB	-49dB
S/N Radio	60 dB	65 dB	67 dB
Impedancia	2.2 kΩ	200Ω	50 Ω
Rango de voltaje	1.5 – 10 V	1.5 V	11V – 52V
Diámetro	9.70 mm	2.5 cm	12.2 mm
Altura	4.70 mm	7.6 cm	37.8 mm
Costo	\$1.200	\$24.900	\$1.225.990

Fuente <https://www.abcelectronica.net/productos/audio/microfono/>,
https://www.audio-technica.com/cms/resource_library/literature/69607746d48985d1/at_eu_products_2007_spain.pdf

Para la selección del micrófono, se tomó en cuenta los parámetros de la tabla 4, en los cuales se puede observar que el micrófono Electret tiene un valor de sensibilidad cercano al de los otros dos micrófonos analizados, lo cual lo presenta como el mejor en relación costo beneficio, adicionalmente cabe resaltar que es el micrófono con las menores dimensiones entre los dispositivos evaluados, por lo cual el micrófono Electret que se observa en la figura 7 es seleccionado para su implementación en el prototipo.

Figura 7. Micrófono Electret



Fuente <https://www.abcelectronica.net/productos/audio/microfono/>

8.5 DISEÑO DE LA GARGANTILLA

Para realizar la gargantilla se requiere un material flexible, que no afecte el proceso de deglución, ni presión el cuello, razón por la cual se escogió como material caucho de ropa que es elástico y permite el movimiento natural del cuello, a continuación, se realizó la toma de medidas del cuello del paciente para realizar la gargantilla, las medidas se presentan en la tabla 5.

Tabla 5. Medidas de la gargantilla

Medidas	
Largo	52cm
Ancho	3cm
Distancia de los micrófonos al borde de la gargantilla	22.5cm

Distancia entre micrófonos	7cm
----------------------------	-----

Fuente Propia

El resultado del diseño con la medida de la tabla 5 se observa en la figura 8, las medidas fueron tomadas a la altura de la manzana de Adan debido a recomendaciones del Otorrinolaringólogo, además de ser el área con mayor claridad en la toma de muestras con el paciente.

Figura 8. Diseño de la gargantilla

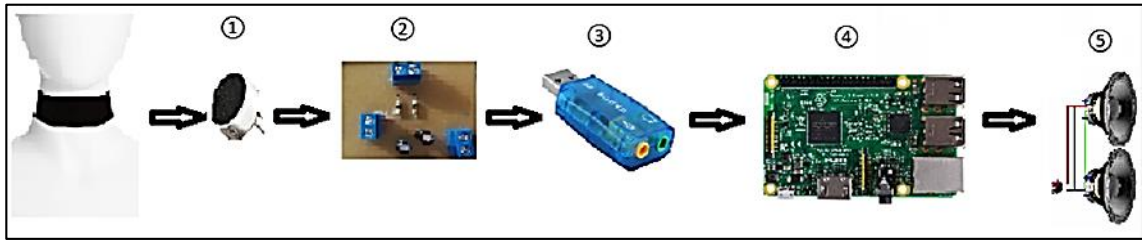


Fuente Propia

8.6 DISEÑO DE PROTOTIPO

A partir del diagnóstico del paciente se diseña un prototipo para captar las señales de las cuerdas vocales de manera no invasiva, de tal forma que no tenga una mayor afectación al paciente, en el momento de realizar la captación de la señal producía por el paciente. El diseño planteado para cumplir tal fin se observa en la figura 9, el cuál consta de 5 elementos que se listan en la tabla 6.

Figura 9. Diagrama prototipo de captación de audio



Fuente Propia

Tabla 6. Componentes utilizados en el prototipo.

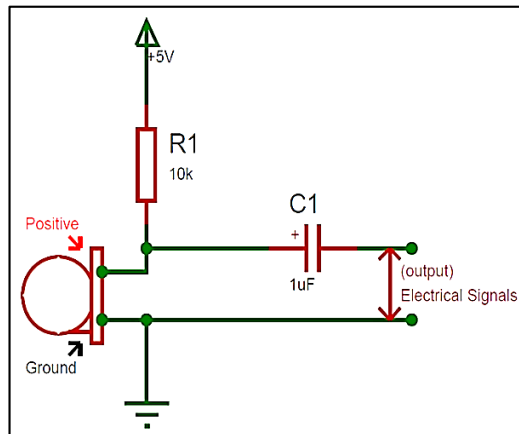
Componente	Elemento	Cantidad
1	Micrófono Electret	2
2	Resistencia	2
2	Condensador	2
2	Jack 3.5	2
2	Conector AK500	3
3	Tarjeta de audio USB	1
4	Raspberry Pi 3B+	1
4	Memoria microSD 64 GB	1
5	Amplificador de audio	1
5	Parlante	2

Fuente Propia

El componente 1 es un micrófono “Electret” con el cual se llevó a cabo la captación de las vibraciones de las cuerdas bucales.

Para el funcionamiento de dicho dispositivo, es necesario incorporar un pequeño circuito con una alimentación entre 2 V y 9, ya que el encapsulado internamente presenta una etapa previa de amplificación, ya que de no contar con esta previa amplificación la salida que entrega el micrófono sería muy pequeña y difícil de manejar, esta alimentación se puede apreciar en la figura 10.

Figura 10. Esquemático PCB



Fuente <https://components101.com/sites/default/files/inline-images/Circuit-using-Electret-Condenser-Microphone.png>

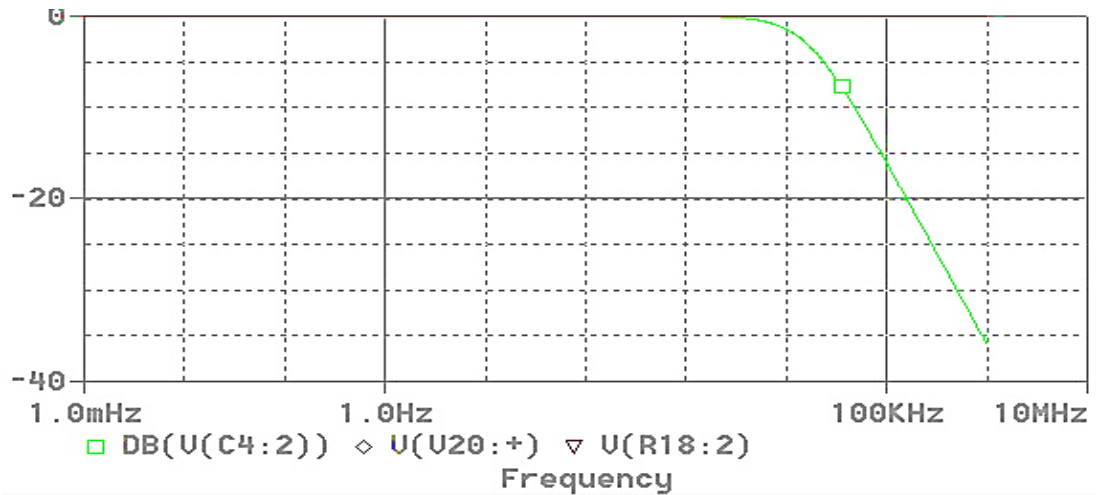
El componente 2 consta de 2 filtros pasa bajos de primer orden, cada uno está compuesto por una resistencia de 10Ω con la finalidad de limitar la corriente que fluye a través del micrófono y un condensador de $1\mu\text{F}$ que limita las altas frecuencias para la eliminación de ruido.

La ecuación 2 da una explicación más detallada de los parámetros que se tomaron en cuenta para el diseño del filtro.

$$R = \frac{1}{2 * \pi * C * F_c} \quad \text{Ecuación 2}$$

Donde (F_c) es la frecuencia de corte, la cual se fijó en 16kHz , se eligió un valor de condensador (C) de $1\mu\text{F}$, para de esta manera obtener un valor de resistencia (R) de 10Ω dando como resultado la gráfica que se observa en la figura 11.

Figura 11. Resultado simulación filtro pasa bajos



Fuente Propia

Al realizar la medición de la señal en -3dB se observa que la frecuencia de corte se encuentra aproximadamente en los 16kHz tal como se aprecia en la figura 12.

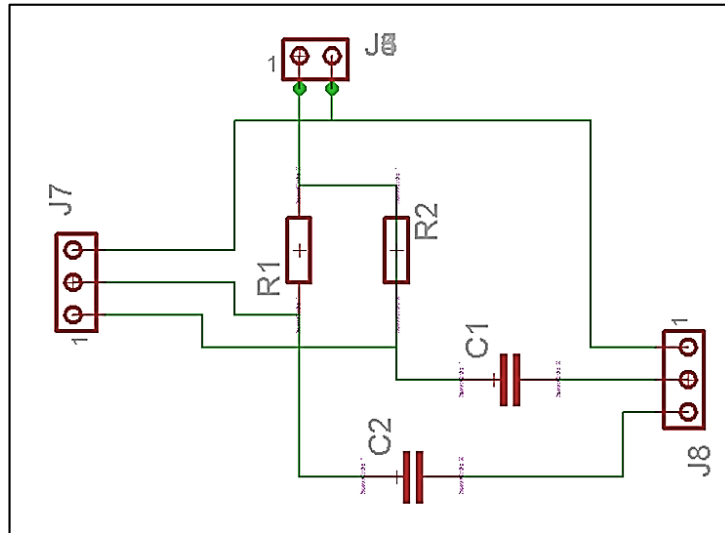
Figura 12. Medición de la señal en -3db

Trace Color	Trace Name	Y1	Y2	Y1 - Y2	Y1(Cursor1) - Y2(Cursor2)	-69.324m			
	X Values	15.909K	15.654K	254.533	Y1 - Y1(Cursor1)	Y2 - Y2(Cursor2)	Max Y	Min Y	Avg Y
CURSOR 1,2	DB(V(C4:2))	-3.0087	-2.9393	-69.324m	0.000	0.000	-2.9393	-3.0087	-2.9740

Fuente Propia

En la figura 13 se presenta una descripción más detallada del esquema utilizado para el diseño de la placa de circuito integrado o PCB por su sigla en inglés (Printed Circuit Board) en la alimentación de los micrófonos.

Figura 13. Esquema de filtro implementado.



Fuente Propia

El componente 3 que se observa en la figura 9 es un adaptador de audio USB, de entrada y salida de audio que transforma señales analógicas a digitales, permitiendo al microprocesador dedicar más recursos de este en tareas más complejas, este dispositivo se puede apreciar en la figura 14.

Figura 14. Tarjeta de audio USB.



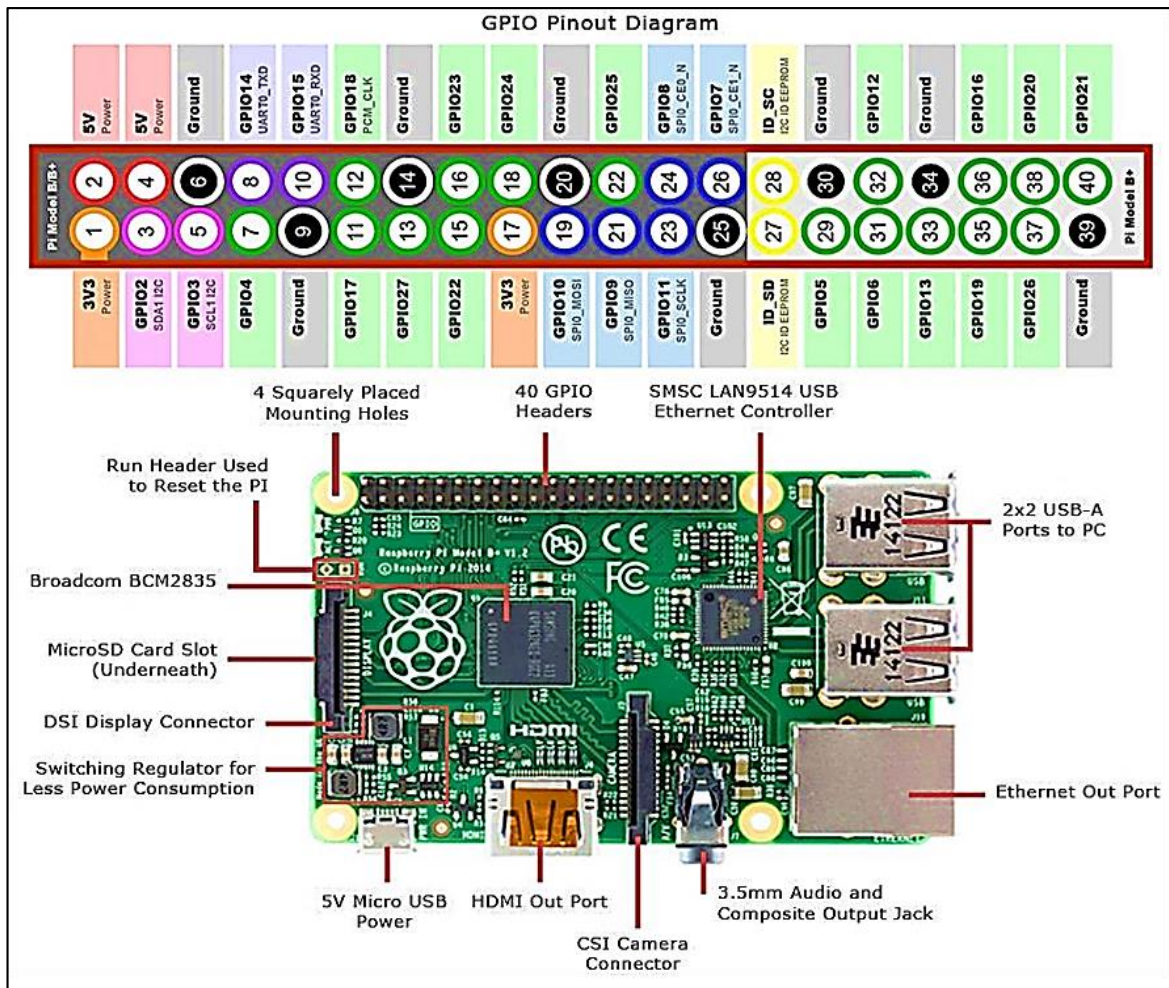
Fuente Propia

El componente 4 es una tarjeta de circuito impreso que cuenta con un sistema embebido conocido como Raspberry Pi 3B+, este dispositivo fue utilizado por su capacidad de procesamiento de información, además de contar con la posibilidad de incrementar el almacenamiento de información del dispositivo en caso de ser necesario, particularmente para el trabajo desarrollado fue utilizada una memoria

microSD “dispositivo digital seguro o Secure Digital por su sigla en inglés”, cuya capacidad es de 64 GB.

Finalmente, esta tarjeta de desarrollo está compuesta por diferentes módulos que son especificados en la figura 15.

Figura 15. Tarjeta Raspberry Pi 3B+.



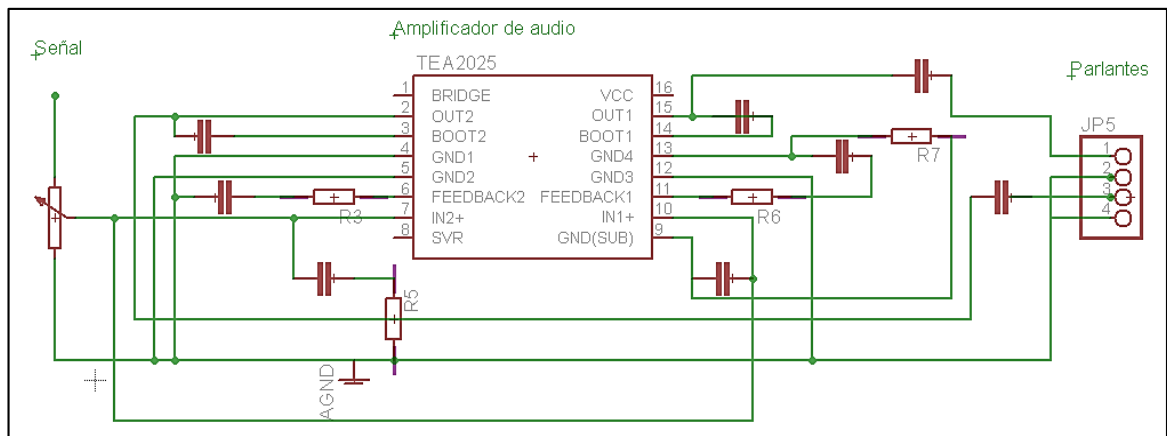
Fuente https://3.bp.blogspot.com/-xYpPeD5bht4/WDGhmkeyaNI/AAAAAAAAABGM/sCLEer-ndJwhG8OJS2s3CaOtIsY2WQMBACLcB/s1600/pi_board_pinout.jpg

Los módulos con mayor utilidad para el procesamiento de la señal de voz captada por el sensor son los que se fundamentan en el protocolo de comunicación de bus serial universal o USB por sus siglas en inglés (Universal Serial Bus), a partir de los

cuales es posible integrar la tarjeta de audio (componente 3) para realizar el debido procesamiento de los datos obtenidos.

Para el funcionamiento de los parlantes que se encuentran en el componente 5 de la figura 9 es necesaria implementar una etapa de amplificación, el elemento utilizado es un circuito integrado TEA2025B con los capacitores y resistencias necesarias para su funcionamiento, este dispositivo cuenta con 2 canales y un potenciómetro para graduar el valor de voltaje de entrada de 3 a 15 generando una ganancia mínima de 43 dB y una ganancia máxima de 47dB, en la figura 16 se puede apreciar el esquemático de este integrado.

Figura 16. Amplificado de Audio TEA2025B



Fuente Propia

Finalmente, el componente 5 que se ve en la figura 17, son los parlantes que reproducen la voz artificial como señal de salida de la tarjeta de desarrollo Raspberry Pi 3B+ (componente 4).

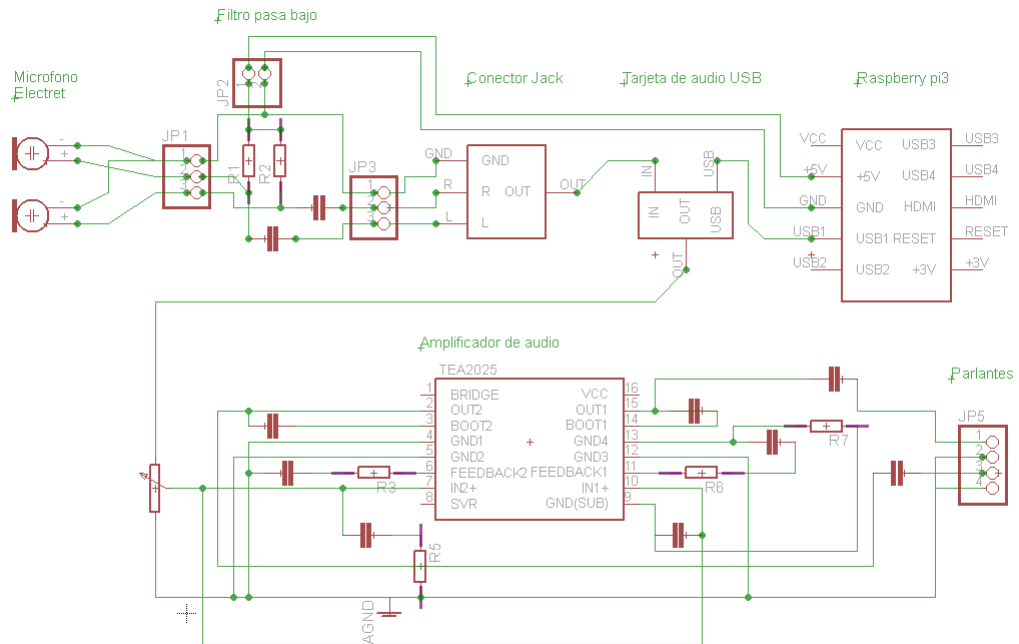
Figura 17. Parlante.



Fuente Propia

Una vez terminada la etapa de diseño se procede a realizar el plano del prototipo con todos los elementos que lo conforman nombrados en la tabla 6, obteniendo como resultado el esquemático presentado en la figura 18, el cual se procede a implementar.

Figura 18. Esquemático del prototipo



Fuente Propia

8.7 ANÁLISIS Y PROCESAMIENTO

8.7.1 Análisis de muestreo. Mediante el prototipo anteriormente presentado, se realiza la captación de las señales producidas por el paciente, proceso en el cual se evidencia la necesidad de la persona en tomar como guía los gestos efectuados por un individuo sano al momento de pronunciar una vocal, intentando reproducir la gesticulación y por consiguiente, mejorar la calidad de la señal producida.

Sin embargo, debido a la dificultad que presenta el paciente en el movimiento de los músculos de la boca que intervienen en el proceso de fonación, se produce una gran variación en los sonidos producidos, lo cual evidencia la necesidad de una base de datos que permita tener una cantidad de muestras considerables para disminuir el porcentaje de error que pueda generar dicha variedad en el sistema, en la tabla 7 se presenta el número de mediciones realizadas al paciente.

Tabla 7. Cantidad de muestras por vocal.

Vocal	Numero de muestras
A	87
E	91
I	41
O	78
U	78

Fuente Propia

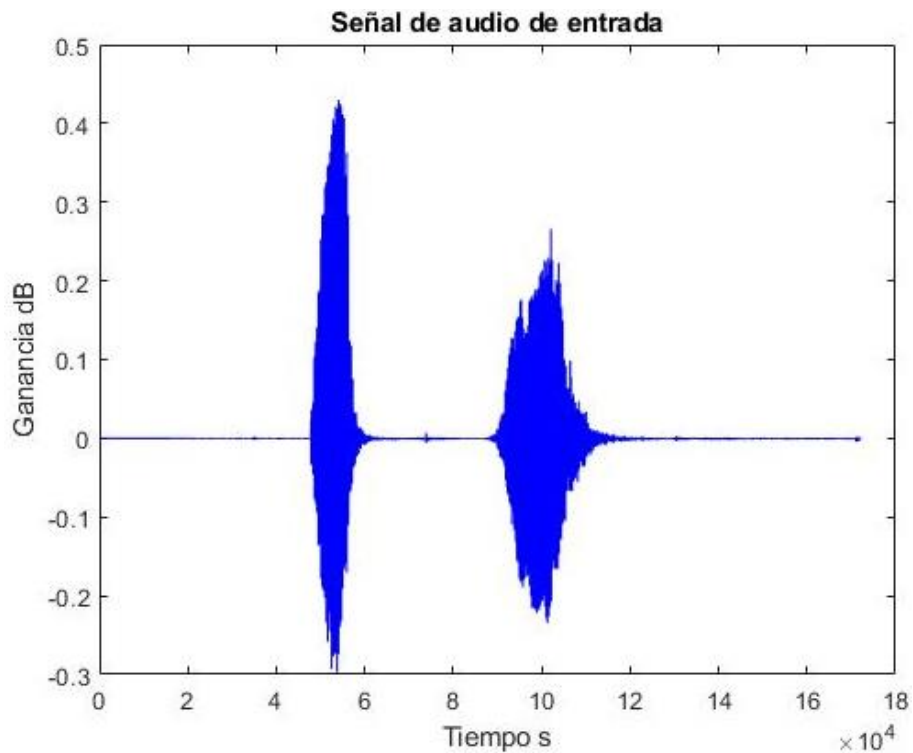
8.7.2 Procesamiento de audio. El análisis de la señal de voz inicia con una caracterización de la señal, esto es realizado con el fin de encontrar las diferencias que presenta la señal producida por una persona sana a la producida por el paciente. Lo anterior es desarrollado mediante el análisis de gráficas en tiempo y frecuencia generadas en el software MATLAB, lo cual permite determinar los posibles métodos de solución al problema que se desea resolver.

El proceso inicia con la lectura de un archivo de audio en el cual se encuentra la voz de una persona sana seguida por la voz del paciente, esto es realizado para conocer información relevante como la frecuencia de muestreo, la duración del audio, y las ganancias producidas por las voces, el resultado es presentado en la figura 19.

A partir de la información contenida en los archivos de audio se procede a iniciar la extracción de señal de paciente y de individuo sano, por esto se normaliza la señal

y se trabaja con los valores de ganancia positivos, dado que los valores negativos son reflejo de estos y por ende solo es necesario tomar un paquete de valores, para lograr lo anterior se aplica la ecuación 3

Figura 19. Señal original.



Fuente Propia

$$y2(i) = \frac{\sum_{i=1}^{N_f-1} y1(i)}{N_f/2} \quad \text{Ecuación 3}$$

Donde y2 es el resultado de la operación, Nf es un valor de 1000 muestras, y1 es el valor absoluto de las muestras de audio obtenidas del paciente, la ecuación 3 es implementada en MATLAB según la descripción del código 1:

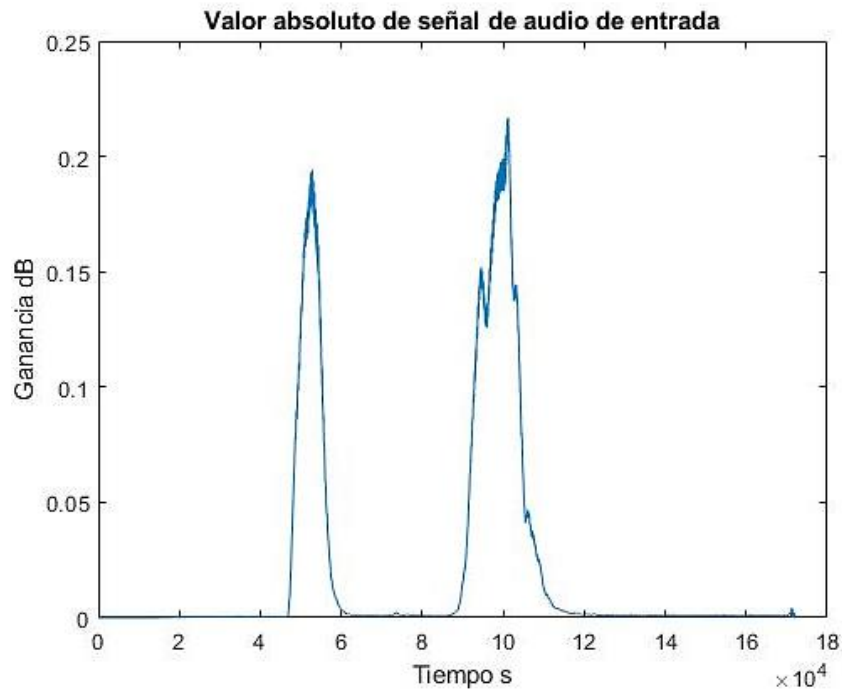
Código 1. Implementación en código de la ecuación 2

```
for i=1:N-Nf
y2(i)=sum(y1(i:i+Nf-1))/500;
end
```

Fuente Propia

Como se observa en el código 1 esta sumatoria se repite N veces, siendo N el tamaño del vector que almacena la señal muestreada.

Figura 20. Señal reajustada con valores positivos.

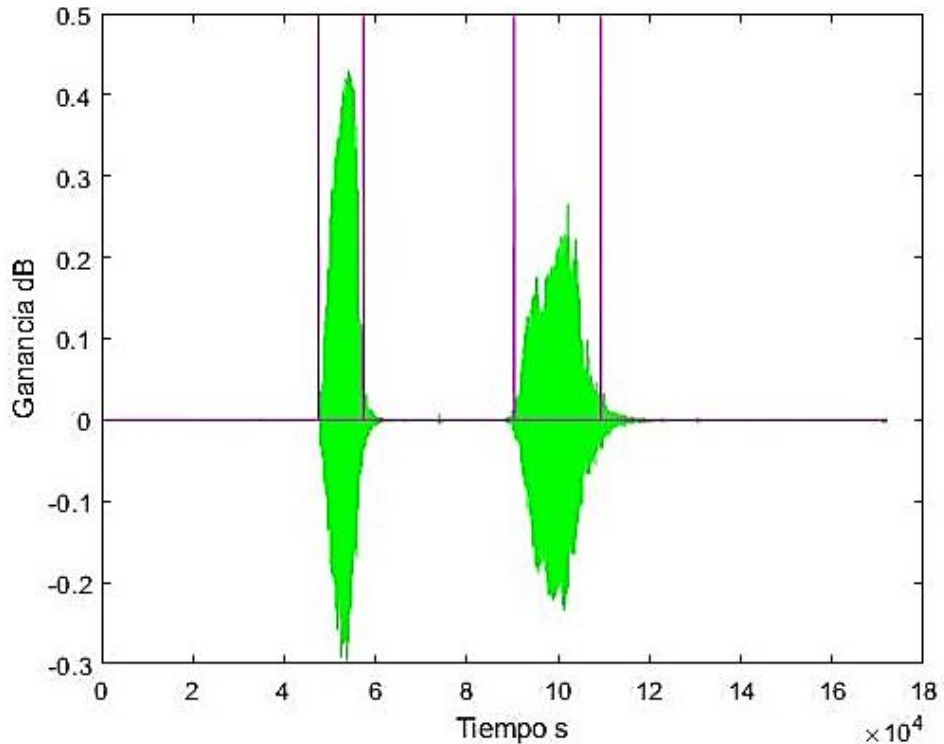


Fuente Propia

La figura 20 presenta el resultado obtenido al implementar el código 1 donde se observan solo los valores positivos, además se reduce la afectación del ruido originado en el medio de captación, debido a que los valores de ganancias de este son bajos comparados con los valores de ganancia de la señal de voz. Adicional a lo anteriormente expuesto la figura 20 permite identificar las secciones de señal donde está la información más relevante para ser procesada como se observa en la figura 21.

A partir de la figura 21 es posible identificar diferencias entre los valores de ganancia para la persona sana de referencia (primer pulso de voz) y el paciente (segundo pulso de voz), con ello es posible iniciar la extracción de la señal del paciente a partir de las muestras de audio, ya que es necesario encontrar el valor de ganancia asociado al paciente, y para lograr dicho objetivo se procesa la señal en MATLAB según el código 2.

Figura 21. Delimitando señales de voz dentro de los audios.



Fuente Propia

Código 2. Hallando las señales de voz dentro de la muestra de audio.

```
ciclos=0
aux=0.02
while(ciclos==0)

    y3=(y2>aux);
    y4=abs(y3(1:length(y3)-1)-y3(2:length(y3)));
    clip=find(y4)
    yJ=y(clip(length(clip)-1):clip(length(clip)))
    fm1=abs(fft(yJ));
    fm2=(fm1(1:floor(length(fm1)/2)))
    ciclos=floor(length(yJ)/2048)
```

```
if(ciclos==0)
    aux=aux+0.01
end
end
```

Fuente Propia

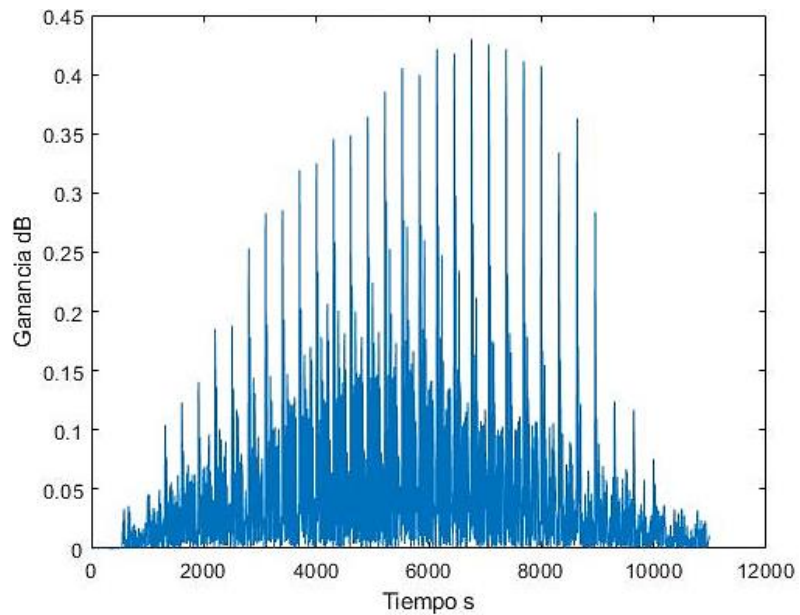
El código 2 cuenta con diferentes variables según el caso:

- ✓ y2 el arreglo de la señal después de procesarlo con la ecuación 1.
- ✓ aux es un auxiliar que permite comparar las ganancias.
- ✓ y3 es la variable que almacena todos los datos que son mayores que aux.
- ✓ y4 es la resta de componentes consecutivos presentes en el arreglo y3.
- ✓ yJ es el vector que almacena la extracción del sonido producido por el paciente.
- ✓ Clip almacena los mayores valores del arreglo y4.
- ✓ Ciclos determina si la cantidad de datos presentes en yJ son suficientes para realizar la transformada de la señal respectiva.

El valor de ganancia se encuentra mediante la variable ciclos, debido a que esta variable va a ser cero siempre que la cantidad de datos abstraídos en el vector yJ, no sean suficientes para realizar una transformada de Fourier.

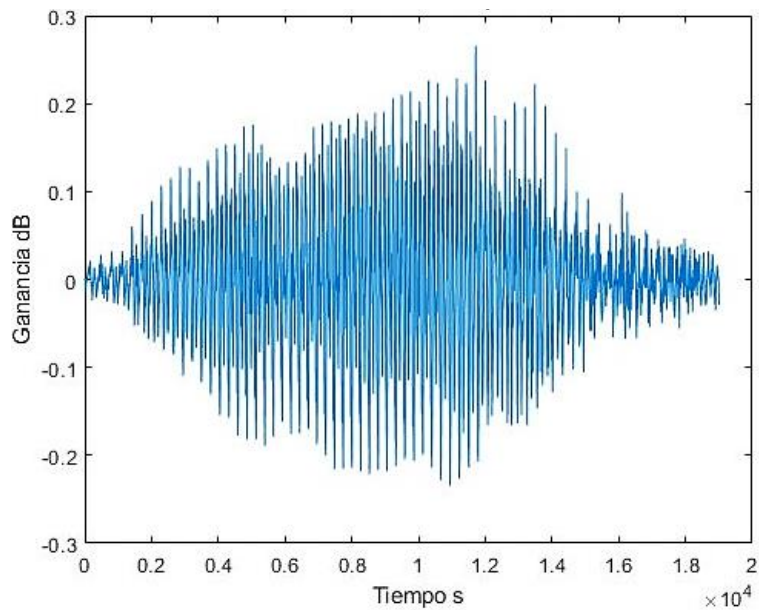
Se procede a graficar el resultado generado en el código 2 para la persona sana (ver figura 22) obteniendo un recorte de información de la señal original y para el paciente obteniendo la señal de la figura 23, donde visualmente se observa que, en la señal producida por una persona sana, la variación de los picos más altos produce una forma de onda más definida que la producida por el paciente; que a su vez tiene más ruido presente en la señal.

Figura 22. Recorte de la señal producida por una persona sana.



Fuente Propia

Figura 23. Señal producida por el paciente.

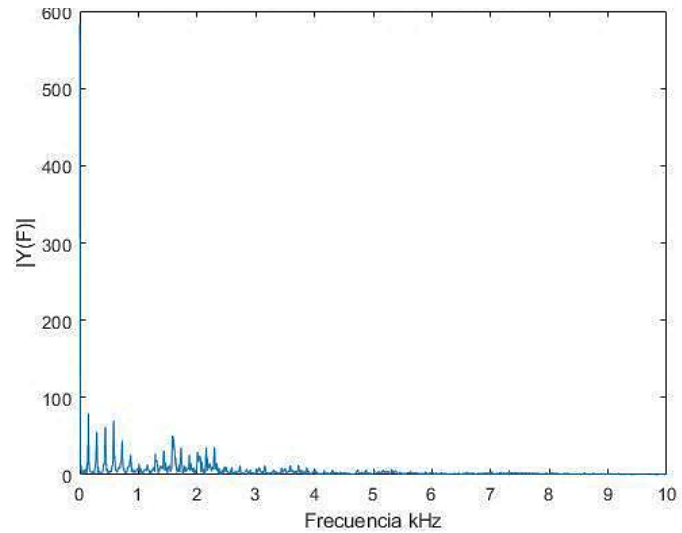


Fuente Propia

A las señales obtenidas se le realiza una transformada de Fourier, para analizar el dominio de la frecuencia determinando las secciones donde se encuentran los armónicos de mayor relevancia para la determinación de las formantes (característica fundamental para identificación de vocales), por ello se realizó la

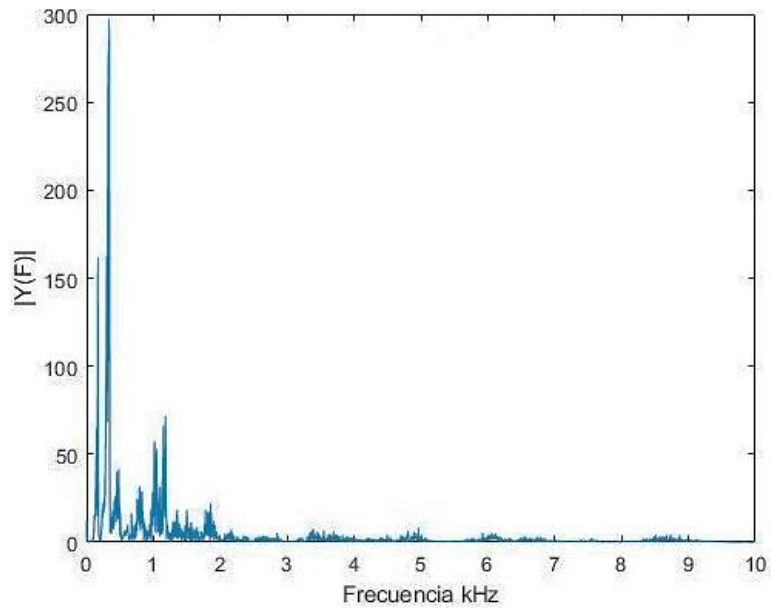
transformada obteniendo como resultado la figura 24 siendo para la persona sana y la figura 25 para el paciente.

Figura 24. Transformada de Fourier de la señal de persona sana.



Fuente Propia

Figura 25. Transformada de Fourier de la señal de voz del paciente.



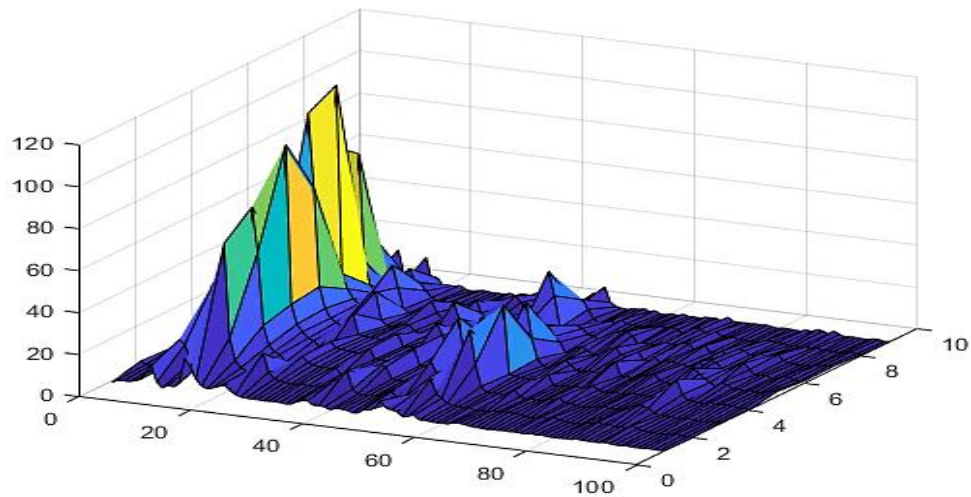
Fuente Propia

Las características que diferencian las dos figuras de las transformadas de Fourier, son las variaciones de los picos de los valores de $|y(f)|$, siendo mayor en la señal producida por el paciente, lo cual sugiere limitaciones en el control de los movimientos de los músculos de la boca a voluntad, induciendo un aumento en el valor de amplitud de 166% aproximadamente, lo cual podría producir errores a tener en cuenta en la caracterización de la señal.

Para observar en detalle la señal producida por el paciente, se procede a realizar la transformada de Fourier en 3d vista en la figura 26, a partir de la cual se observan picos máximos en frecuencias donde se concentra la información.

En el dominio de la frecuencia, la detección de componentes fundamentales permite analizar frecuencias con mayor intensidad y así encontrar armónicos más cercanos a la frecuencia fundamental.

Figura 26. Transformada de Fourier del paciente en 3d.



Fuente Propia

Las vocales se producen mediante ondas acústicas, cada una de estas tiene su espectro característico: la A y la U tienen fuerte la componente fundamental y el tercer armónico, mientras que el segundo y cuarto son débiles; la E y la O, tienen el segundo y cuarto armónico fuertes, mientras la fundamental y el tercer armónico débil, por último, la vocal I tiene sus primeros armónicos débiles y el quinto y sexto fuerte.

8.7.3 Método de solución del proyecto. Los datos obtenidos por medición, permiten la creación de una base de datos que contiene 300 muestras, la cual es procesada para crear un archivo de extensión cvs (comma-separated values), que posteriormente es manipulado por una red neuronal, la distribución de las muestras se observa en la tabla 8.

Tabla 8. Cantidad de muestras por vocal.

Vocal	Cantidad de muestras para entrenar
A	60
E	60
I	60
O	60
U	60

Fuente Propia

La selección de redes neuronales es debida a que son algoritmos matemáticos que permiten procesar datos de una manera más versátil; puesto que, facilitan la búsqueda de patrones. Esto es muy útil para el presente proyecto, al considerar el diagnóstico severo del paciente en la dificultad de su fonación, además que no es evidente la similitud entre las señales producidas por una misma vocal; por lo que, el uso de redes neuronales permitirá identificar determinadas frecuencias en las que se supone se encuentra la información de las vocales producidas por el paciente.

La red neuronal es implementada a partir del lenguaje Python, esta utiliza una red de clasificación que tiene como objetivo analizar la señal obtenida del paciente y procesarlo para encontrar patrones que le permita relacionar la señal producida con algunas de las presentes en la base de datos.

Adicional a lo anterior, se incluyen filtros adaptativos, los cuales se aplican en sistemas variables, porque presentan características autoajustables, capaces de eliminar el ruido en las señales mediante un ajuste en su respuesta en frecuencia.

Para calcular dichas frecuencias, es necesario tener en cuenta los coeficientes proporcionados por las redes neuronales; dichos coeficientes son resultado del entrenamiento de la red neuronal y dependen de los valores de entrada de la señal.

8.7.4 Diseño de la red neuronal de entrenamiento. Para construir el dataset de la red neuronal es necesario almacenar en un archivo con formato cvs, una matriz con la información de las muestras, el número de la muestra y la señal procesada del paciente, adicionalmente en un archivo de extensión .dat (data) se almacena un vector con cada coeficiente asignado a cada vocal como se observa en la Tabla 9.

Tabla 9. Pesos característicos del target para la red.

Vocal	Coeficiente
A	0
E	1
I	2
O	3
U	4

Fuente Propia

A continuación, se cargan los archivos con extensión (.cvs) y (.dat) en la red neuronal, con la finalidad de asignar dicha información de cada archivo a una variable. Python permite el uso de librerías con modelos preestablecidos de diferentes tipos de red, por dicha razón se utilizó la biblioteca keras. presente en el código 3.

Código 3. Librerías de la red neuronal.

```
from sklearn.model_selection import train_test_split
from keras.utils import to_categorical
from keras.models import Sequential
from keras.layers import Dense, Flatten
from keras.models import model_from_json
```

Fuente Propia

La codificación para la red neuronal seleccionada contenida en la librería keras es presentada en el código 4.

Código 4. Creación del modelo de red neuronal.

```
classifier.add(Dense(output_dim=11, init='uniform',activation='relu',input_dim=50000 ))
classifier.add(Dense(output_dim=9, init='uniform',activation='relu' ))
classifier.add(Dense(output_dim=7, init='uniform',activation='relu' ))
classifier.add(Dense(output_dim=5, init='uniform',activation='softmax'))
```

Fuente Propia

Seleccionado el tipo de modelo de la red neuronal se procede a realizar la configuración de los parámetros de aprendizaje como se observa en el código 4, en el cual el Classifier.add representa el constructor de una capa de la red clasificadora, cuya repetición (cuatro veces) indica el número de capas de la red neuronal cuatro para este caso.

Las capas tienen cada una como parámetros de entrada la cantidad de neuronas, el tipo de distribución, y la función de activación, que en este caso es seleccionada una función relu (unidad rectificadora lineal), representada con la siguiente expresión matemática (ecuación 4):

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{para } x < 0 \\ x & \text{para } x \geq 0 \end{cases} \quad \text{Ecuación 4}$$

Siendo x el valor de amplitud de la señal donde la función $\max(0, x)$ define 0 para los valores negativos y mantiene los valores positivos fijos en su respectiva posición, permitiendo que el aprendizaje de la red solo tome en cuenta los coeficientes positivos mejorando el reconocimiento.

Para la última capa o capa de salida, se realiza un cambio en la función de activación para utilizar la función softmax, la cual informa la probabilidad que presenta la entrada de los datos de pertenecer a una de las clases configuradas, en este caso determina la probabilidad del audio de entrada de pertenecer a alguna de las vocales.

El optimizador de la red diseñada es Adam como se evidencia en el código 5, debido a que, adapta el radio de aprendizaje de acuerdo con la distribución de los parámetros de aprendizaje, dado que tener datos con mayor dispersión incrementará el radio de aprendizaje. Además, para la estimación de error de la red se seleccionó Categorical Cross-Entropy debido a su precisión para variables categóricas, su representación matemática se observa en la ecuación 5

Código 5. Optimizador de la red neuronal.

```
classifier.compile('adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

Fuente Propia

$$\mathcal{L}(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log(p_{ij})$$

Ecuación 5

Donde n es el tamaño de la muestra de audio, m es el número de categorías que para el presente trabajo es 5 por la cantidad de vocales, y_{ij} es el vector de aprendizaje y p_{ij} es el modelo utilizado para el aprendizaje el cuál es la probabilidad pronosticada por el modelo. Finalmente, en el código la métrica es realizada por el uso del modelo llamado “accuracy”, elemento encargado de averiguar la cantidad de aciertos donde el modelo predice el resultado de forma correcta.

La codificación del entrenamiento de la red neuronal previamente creada es presentada en el código 6, en el cual se observan los parámetros X_{Train} , que representa la entrada de los datos, $target_{train}$ que es un vector donde se almacena la representación del comportamiento deseado, el $batch_size$ es el tamaño del lote o la cantidad de muestras de audio que se envían a la red en cada entrenamiento, y por ultimo $epoch$ es el número de iteraciones de la red por lote para calcular los coeficientes de aprendizaje de la red, la cual se conoce como épocas.

Código 6. Método de entrenamiento

```
classifier.fit(X_Train,target_train, batch_size=40, nb_epoch=200)
```

Fuente Propia

El modelo de almacenamiento de datos es determinado por json desde la librería de keras, el cual crea y guarda el modelo de las predicciones de la red neuronal. Esta librería es utilizada debido a que json es un formato de texto sencillo para el intercambio de datos, dicho elemento es configurado en el código 7.

Código 7. Modelo de almacenamiento de datos.

```
from keras.models import model_from_json
redvocales_json = classifier.to_json()
with open("red_vocalesp.json", "w") as json_file:
    classifier.save_weights("red_vocalesp.h5")
print("Saved red to disk")
```

Fuente Propia

La librería keras permite almacenar el modelo de la red neuronal mediante el formato json que es un formato de archivos para la descripción de datos jerárquicos, los pesos calculados en el modelo se almacenan en un archivo con extensión h5 cuyo formato se caracteriza por una cuadrícula ideal para almacenar matrices multidimensionales de números, el cual se conoce como hdf5.

En el programa principal se carga una señal de audio que contiene una vocal a clasificar, se procesa y posteriormente se carga el modelo de predicción de la red neuronal el cual se encuentra en el archivo json, así como los coeficientes de aprendizaje del archivo h5 como se observa en el código 8.

Código 8. Carga del modelo de predicción y los pesos

```
json_file = open('red_vocalesp.json', 'r')
loaded_model_json = json_file.read()
json_file.close()
loaded_model = model_from_json(loaded_model_json)
# load weights into new model
loaded_model.load_weights("red_vocalesp.h5")
print("Loaded model from disk")
```

Fuente Propia

Luego de cargar la señal y el modelo de la red se da inicio al proceso de predicción de la red tras ingresar la señal de voz a la red neuronal cómo se observa en el código 9.

Código 9. Carga de la señal de entrada en el modelo de predicción.

```
entrada[0,0:tamm1]=Yj[0:tamm1]
result = loaded_model.predict(entrada)
result=np.array(result)
print(result)
```

Fuente Propia

El modelo de predicción da como resultado un vector que contiene una posición por cada clase, cinco en total para el presente proyecto, una por cada vocal. El vector cuenta con diferentes datos siendo el de mayor magnitud el que la red neuronal predice como respuesta, es importante recordar que las posiciones de izquierda a derecha son las vocales a, e, i, o, u, como se evidencia en el código 10.

Código 10. Resultado de la clasificación de la red

```
[1. 0. 0. 0. 0.]
La vocal ingresada es la A
```

Fuente Propia

Una vez identificada la vocal mediante la librería de texto a voz llamada pyttsx3, cuya implementación en código se observa en el código 11

Código 11. Implementación de texto a voz de la librería pyttsx3

```
engine = pyttsx3.init()
engine.setProperty('volume',5.0)
engine.setProperty('rate',230)
engine.setProperty('voice','spanish')
engine.say('A')
engine.runAndWait()
```

Fuente Propia

La implementación del código 11 envía la señal a la tarjeta de audio que está conectada a una etapa de amplificaron debido a que la señal de salida presenta una

ganancia muy baja por la que se amplifica y se envía a los parlantes para reproducirla.

9 ANÁLISIS DE RESULTADOS

Las pruebas realizadas con la red neuronal entrenada produjeron los resultados relacionados en la Tabla 10.

Tabla 10. Pruebas de la red neuronal con las muestras del paciente.

Vocal	Número de pruebas	Porcentaje de aciertos
A	150	46%
E	140	32%
I	75	29%
O	125	35%
U	130	43%

Fuente Propia

Los resultados obtenidos se deben a que la red neuronal no es capaz de diferenciar efectivamente las vocales, debido que los armónicos de las señales que produce el paciente están desplazados hacia la derecha, esto es importante, porque los armónicos son las formas de vibración de las cuerdas bucales, la forma de vibración que presenta una mayor amplitud se llama frecuencia fundamental y es la que contiene más energía acústica, pero debido al diagnóstico del paciente, la diferencia entre la frecuencia fundamental y los otros armónicos no es perceptible.

Para realizar el análisis de los resultados obtenidos por la red se realiza un estudio del espectro de las señales, el cual nos permite comprender como el diagnóstico del paciente afecta la producción de los sonidos de los fonemas, es importante resaltar nuevamente que las formantes en las personas sanas cuando entran en contacto con órganos como la lengua o la faringe, amplifican o atenúan algunas frecuencias que determinan el fonema producido, además producen una resonancia que amplifica la amplitud de ciertos rangos de frecuencias.

Debido al diagnóstico del paciente la producción de fonemas se ve alterada debido a la disminución de movimientos de órganos como la lengua y la faringe lo que no permite modificar ciertos armónicos impidiendo que los fonemas producidos tengan diferentes características, dificultando el proceso de reconocimiento.

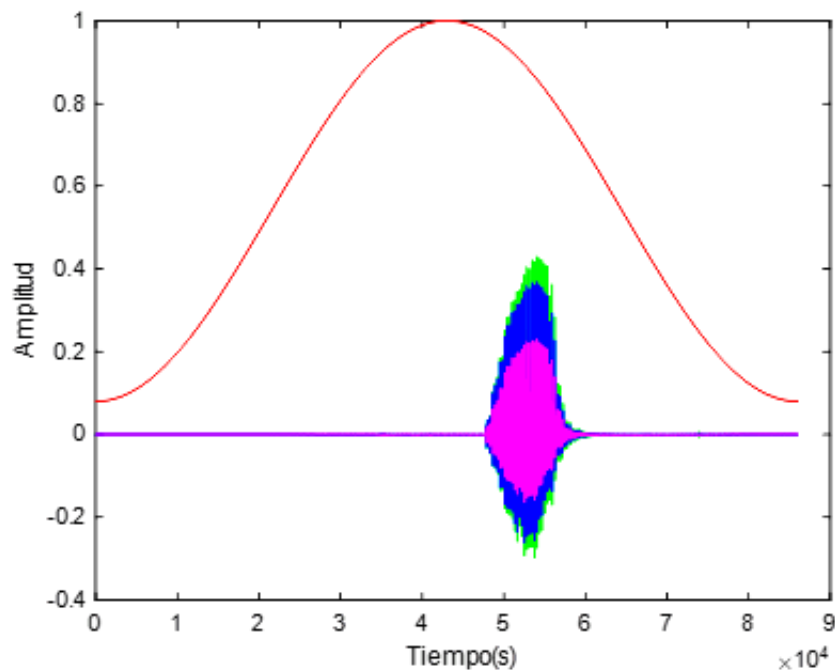
De acuerdo con lo expuesto anteriormente se procede a realizar la comparación de las formantes presentes en la señal de un individuo sano y el paciente para conocer

que tanto son alteradas las formantes debido a la afección del paciente, lo cual requiere la implementación de un filtro pasa-bajos para posteriormente realizar una ventana de hamming.

Una ventana diseñada en el dominio del tiempo aplicada a una señal es comparable al proceso de convolución en el dominio de la frecuencia, adicional dicha ventana tiene tres propiedades las cuales son el ancho del lóbulo principal, un nivel de lóbulos laterales y una pendiente de caída.

La ventana de hamming hace uso de una forma sinusoidal con un amplio pico, pero sus lóbulos laterales no llegan a cero, debido a lo cual se usan para medidas de ruido para mejorar la resolución de frecuencia. A partir de lo anterior se hace uso de los módulos de Matlab, para realizar una ventana de Haming a la señal generada por el individuo sano como se observa en la figura 27.

Figura 27. Ventana de Hamming para señal de una persona sana.



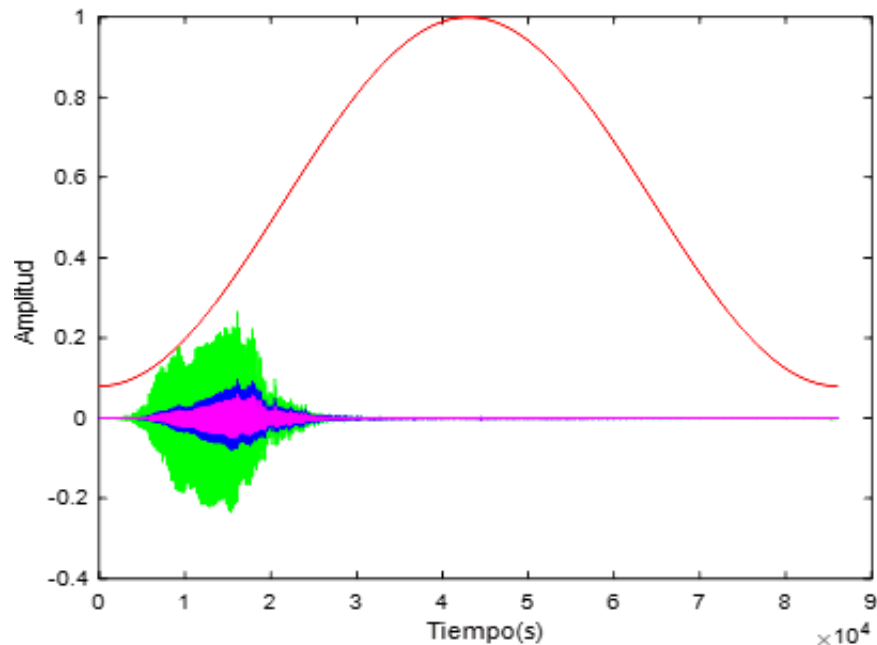
Fuente Propia

La figura 27 permite observar la ventana de Haming de color rojo, la señal inicial de color verde, la señal azul representa la aplicación de la ventana del procesamiento a la señal y por último el color morado es la señal después de aplicarle un filtro pasa

altos, observando que la señal del individuo se encuentra cerca de la zona central del lóbulo principal.

El mismo proceso descrito para un persona sana es realizado para el paciente presentado en la figura 28, donde se puede observar la ventana de Hamming de color rojo, de color verde la señal inicial, la señal azul representa la aplicación de la ventana del procesamiento a la señal y por último el color morado es la señal después de aplicarle un filtro pasa altos, en esta ventana se puede apreciar claramente que la señal no se encuentra en el lóbulo principal, sino que tiende a encontrarse en uno de los lóbulos laterales, lo cual puede deberse a la resonancia acústica producida por la disminución de movimiento de músculos como los labios o la lengua

Figura 28. Ventana de Hamming para señal producida por el paciente



Fuente Propia

A los resultados de las ventanas de Hamming se les implementa el método LPC (Linear Predictive Code), siendo un análisis de series temporales conocido como modelo Autorregresivo (ecuación 6), para permitir el cálculo de los formantes.

$$S_n = -\frac{1}{n} \sum_{k=1}^p a_k S_{n-k} + G U_n \quad \text{Ecuación 6}$$

Donde S_n es una serie temporal dada como una combinación lineal de sus p valores previos, y de una entrada U_n , a_k es el coeficiente de predicción lineal y G la ganancia de la señal. A la ecuación 4 es necesario realizar la transformada de Fourier, a partir de la cual se obtiene la ecuación 7.

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{k=1}^p a_k z^{-k}} \quad \text{Ecuación 7}$$

A partir de la transformada de Fourier es posible conocer los valores de los formantes presentes en las señales de audio, teniendo en cuenta el coeficiente de correlación que indica la cantidad de frecuencias fundamentales que se necesitan, por ello se establece el cálculo del coeficiente de auto correlación en la ecuación 7.

$$R(i) = \sum_{n=-\infty}^{\infty} S_n S_{n+i} \quad \text{Ecuación 7}$$

La implementación de las ecuaciones anteriormente desarrolladas se presenta en el código 11, donde $xJ1$ es el vector que contiene la señal de entrada y el número 16 representa el orden del modelo de predicción lineal.

Código 12. Estimación de coeficientes LPC.

```
A = lpc(xJ1,16);
rts = roots(A);
rts = rts(imag(rts)>=0);
```

Fuente Propia

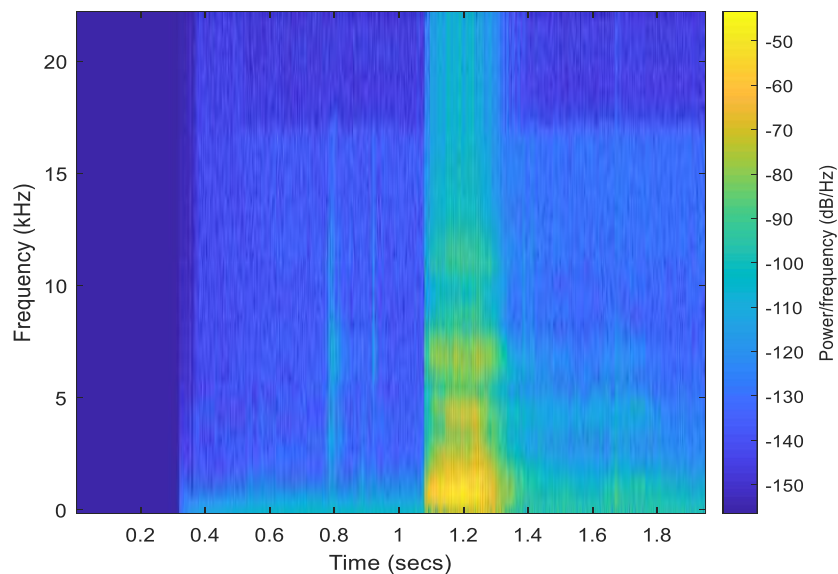
La variable rts contiene los coeficientes de predicción lineal, a partir de los datos ingresados.

El espectrograma permite hacer un análisis visual de los valores en los que se encuentran las frecuencias fundamentales, Gracias a que el espectrograma es una representación de la distribución de energía sonora, es posible tener una visualización clara del sonido ya que esta es de naturaleza espectral, de esta forma es posible tener en cuenta características de la señal que antes no eran percibidas de forma audible.

En la señal producida por el individuo sano como se observa en la figura 29, los valores de ganancia sobre frecuencia se encuentran entre 50 y 70 dB/Hz, ubicados entre 1 y 1.4 segundos, además podemos apreciar una visualización más clara de los formantes, mientras que en el espectrograma de la figura 30 se aprecia que los niveles de ganancia sobre frecuencia se encuentran demasiado cercanos, dificultando la diferenciación de los formantes para el paciente.

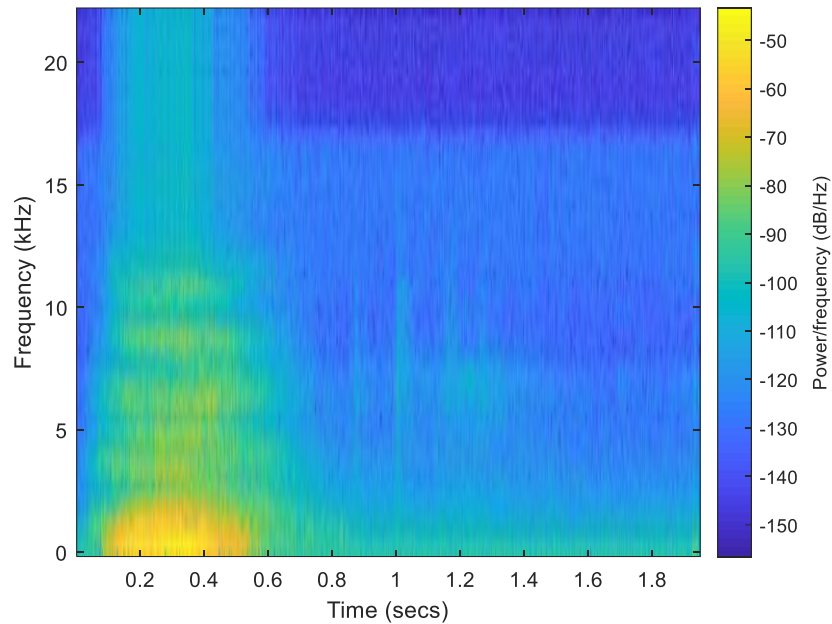
En el espectrograma generado por el paciente se puede apreciar como los valores de ganancia sobre frecuencia son iguales al rango producidos por una persona normal, pero el intervalo de tiempo si se modifica, desplazándose a la derecha.

Figura 29. Espectrograma de la señal producida por una persona sana.



Fuente Propia

Figura 30. Espectrograma de la señal producida por el paciente.



Fuente Propia

Para determinar la incidencia del diagnóstico del paciente en los resultados visualizados en el espectrograma y por ende en la clasificación de la red neuronal, se obtienen 25 muestras de audio para cada vocal, a cada una de las muestras se le halla el valor del primer y segundo formante, los cuales son fundamentales en la clasificación de los fonemas vocales.

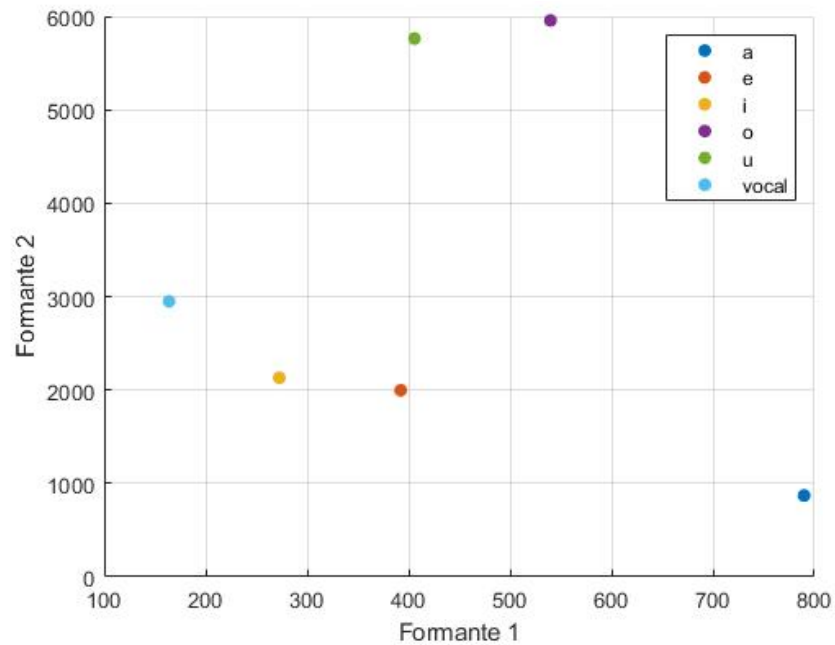
Posteriormente es calculado el promedio de los formantes, los cuales se grafican, por medio de puntos como se ve en figura 31. Cada punto muestra la posición de los cuadrantes característicos para cada vocal según el cálculo obtenido de las formantes.

Al ingresar una nueva entrada de audio del paciente y calculando el primer y segundo formante es posible posicionar el cuadrante característico de la señal producida por el paciente y compararlos con los promedios obtenidos de una persona sana, siendo la entrada de audio del paciente aquella que aparece en la leyenda de la gráfica como “vocal”.

La intervención de los músculos de la vocal en la producción de los fonemas tiene una incidencia en el valor de los formantes, ya que con la sola variación de la posición de las mejillas para aumentar o disminuir la cantidad de aire, afecta la ganancia con la que es producida, además de cambiar la resonancia de la cavidad

bucal. Adicionalmente la posición de la lengua y la epiglotis pueden tener incidencia en la ganancia, o atenuar frecuencias que son indispensables para cambiar el fonema producido.

Figura 31. Diagrama de formantes producidos por el paciente (vocal E).



Fuente Propia.

10 CONCLUSIONES

El desarrollo del dispositivo de captación de señales de audio cuenta con un diseño mínimamente invasivo, lo que permite tener baja o nula incidencia en la salud del paciente, dado su tamaño reducido considerando las etapas de amplificación, de filtrado, y procesamiento robusto de las señales.

A partir de la solución planteada, fue posible seleccionar la tarjeta Raspberry Pi3 la cual cuenta con características suficientemente robustas para procesar la red neuronal descrita en Python, que opera las señales vibratorias producidas por las cuerdas bucales, adicionalmente se implementó un filtro y se usó una tarjeta de audio USB de forma que el hardware reduzca la carga de procesamiento en software.

Las señales de voz en el dominio del tiempo producidas por un paciente con radio necrosis frontoparietal izquierda, es posible evidenciar la afectación de salud en las ganancias presentes en la señal captada por los sensores, además el proceso reiterado de habla y el sobre esfuerzo a la hora de producir sonidos por el paciente ocasiona una fatiga muscular, además el reducido control del paciente del aparato fonador incrementa la afectación de la señal generada.

Las señales de voz en el dominio de la frecuencia producidas por un paciente con radio necrosis frontoparietal izquierda, permiten observar a partir del espectrograma obtenido por la persona afectada como se aprecia en la figura 30, como los valores de ganancia sobre frecuencia son iguales al rango producidos por una persona normal, pero el intervalo de tiempo es modificado, desplazándose a la derecha, dado que la intervención de los músculos de la boca en la producción de los fonemas tiene una incidencia en el valor de los formantes, ya que con la sola variación de la posición de las mejillas para aumentar o disminuir la cantidad de aire, afecta la ganancia con la que es producida, además de cambiar la resonancia de la cavidad bucal.

El prototipo permite generar una señal de audio a partir de una tarjeta de sonido y unos parlantes implementados en el diseño, dicha señal de audio se produce una vez la red neuronal a identificado la vocal de entrada al sistema, el software reproduce el archivo de audio correspondiente a la respuesta entregada.

BIBLIOGRAFÍA

- [1] J. L. F. K. Ishizaka, «Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords,» *The Bell System Technical Journal* , vol. 51, pp. 1233 - 1268, 1972.
- [2] Damian Krzesimowski, «Voice Signal Processing for Patients with Stroke Hospitalisation,» de *Mixed Design of Integrated Circuits & Systems*, 2009. MIXDES '09. MIXDES-16th International Conference, Lodz, Polonia, 2009.
- [3] S.-F. C. C.-C. C. Chien-San Lin, «Microwave Human Vocal Vibration Signal Detection Based on Doppler Radar Technology,» *IEEE Transactions on Microwave Theory and Techniques* , vol. 58, nº 8, pp. 2299 - 2306, 2010.
- [4] N. R. R. H. N. U. V. S. Balaji, «Identification of Predominant Frequencies in a Speech signal Using Modeling of Vocal Chord,» de *Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEeCE)*, 2011 International Conference on, Sivakasi, India, 2011
- [5] R. D. P. C. V. L. W. A. R. Martin Andermann, «Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex,» *NeuroImage*, vol. 158, nº 158, pp. 79-89, 2017.
- [6] B. M. M. Cristian Herrera Hernandez, «virtuniversidad,» [En línea]. Available:
<https://www.virtuniversidad.com/greenstone/collect/ingles/index/assoc/HA SHf01b/45450734.dir/doc.pdf>. [Último acceso: 12 abril 2018].
- [7] A. C. A. V. F. Casassa, «A Phonatory System Simulator for testing purposes of voice-monitoring contact sensors,» de *Instrumentation and Measurement Technology Conference (I2MTC)*, 2017 IEEE International, Turin Italia, 2017.
- [8] P. M. P. V. A. A. F. SusanaVaz-Freitas, «Acoustic analysis of voice signal: Comparison of four applications software,» *Biomedical Signal Processing and Control*, vol. 40, nº 158, pp. 318-323, 2018.
- [9] E. d. r. y. e. d. e. m. d. l. S. A. C. E. Cáncer, «American Cancer Society,» American Cancer Society, 26 octubre 2017. [En línea]. Available: <https://www.cancer.org/es/cancer/tumores-de-encefalo-o-de-medula-espinal/acerca/estadisticas-clave.html>. [Último acceso: 2 mayo 2018].
- [10] Herrera, C., & Begoña, M. (n.d.), «El aparato fonador. Física- Fisiología II.,» [En línea]. Available: <https://www.virtuniversidad.com/greenstone/collect/ingles/index/assoc/HA SHf01b/45450734.dir/doc.pdf>. [Último acceso: 29 09 2019]
- [11] Dra. Begoña Torres, «Anatomía funcional de la voz» [En línea]. Available: <http://www.medicinadelcant.com/cast/1.pdf>. [Último acceso: 29 09 2019]
- [12] musiki, «musiki» [En línea]. Available: <http://musiki.org.ar/Formantes>. [Último acceso: 23 09 2019]

- [13] «psicoactiva,» [En línea]. Available: <https://www.psicoactiva.com/blog/lobulo-parietal-anatomia-funcion/>. [Último acceso: 2019 09 29].
- [14] N. Contemporánea, «Neurocirugía Contemporánea,» 2008. [En línea]. Available: <http://neurocirugiacontemporanea.com/doku.php?id=radionecrosis>. [Último acceso: 29 09 2019].
- [15] «Audio y control digital,» [En línea]. Available: http://catarina.udlap.mx/u_dl_a/tales/documentos/lep/martinez_n_lr/capitulo3.pdf. [Último acceso: 2019 09 29].
- [16] L. d. a. d. f. aplicados, «transformada de fourier,» [En línea]. Available: http://www4.ujaen.es/~jalmira/transformada_fourier_almira.pdf. [Último acceso: 2019 09 2019].
- [17] SCRIBD, «SCRIBD» [En línea]. Available: <https://es.scribd.com/document/358481883/Ventanas-de-Hamming> [Último acceso: 23 09 2019]
- [18] musiki, «musiki» [En línea]. Available: [http://musiki.org.ar/Espectrograma_\(sonograma\)](http://musiki.org.ar/Espectrograma_(sonograma)) [Último acceso: 25 07 2019]
- [19] J. A.-. T. Barrera, «Redes Neuronales,» [En línea]. Available: http://www.cucei.udg.mx/sites/default/files/pdf/toral_barrera_jamie_areli.pdf. [Último acceso: 29 09 2019].
- [20] Sarika Hagde, S. S. (2020, 06 10). *A Survey on Machine Learning Approaches for Automatic*. Retrieved from sci-hub: <https://sci-hub.tw/10.1016/j.jvoice.2018.07.014>

ANEXOS

Julio Cesar Convers autoriza de forma escrita el manejo de su historial médico para la realización de este proyecto como se observa a continuación.



Bogotá, 5 de junio de 2018

Señores
Universidad Santo Tomás
Calle 52 No. 9 – 14
Bogotá D.C.

Asunto: autorización de manejo de información de historial médico.

Respetados señores,

Yo, Julio Cesar Convers con número de ciudadanía 13722684 de B/moya, mediante la presente autorizo el acceso a mi historia clínica y diagnósticos a los estudiantes de Ingeniería Electrónica de la Universidad Santo Tomás sede Bogotá Juan Manuel Silva Rojas, identificado con cedula de ciudadanía 1052400167 de Duitama y Julián Camilo Ortiz, identificado con cedula de ciudadanía 1018488265 de Bogotá, para el desarrollo del proyecto "Diseño e implementación de un modelo funcional de generación de voz por medio de la captación de vibración de cuerdas vocales de personas con discapacidad de comunicación oral", bajo la supervisión de **M.Sc. Oscar Mauricio Gelvez Lizarazo** docente de tiempo completo del grupo de investigación MEM (Modela Electrónica y Monitoreo) de la Universidad Santo Tomás.

FIRMA

Julio Cesar Convers
C.C. 13722684

GLOSARIO

1. **Afasia:** Trastorno del lenguaje que se caracteriza por la incapacidad o dificultad de comunicarse mediante el habla
2. **Agrafia:** Incapacidad, debido a lesión cerebral, para comunicar ideas y los pensamientos por escrito.
3. **Fonema:** Son unidades de análisis lingüístico que esta basada en los sonidos de una lengua.
4. **Formantes:** Es el pico de intensidad en el espectro de un sonido, es la amplitud de onda que se da en una determinada frecuencia.
5. **Radioterapia:** Tratamiento médico de algunas enfermedades, en el cual involucra el uso de todo tipo de rayos, especialmente rayos X.
6. **Ventana de hamming:** Función matemática usada con frecuencia en el análisis y procesamiento de señales para evitar las discontinuidades al principio y al final del bloque analizado.