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



Autores:

Juan Manuel Silva Rojas

Julián Camilo Ortiz Jiménez

Universidad Santo Tomás

Facultad de Ingeniería Electrónica

Programa de Ingeniería Electrónica

Bogotá D.C

2019

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

Autores:

Juan Manuel Silva Rojas

Julián Camilo Ortiz Jiménez

Director: **M.Sc.** Oscar Mauricio Gelvez Lizarazo

Codirector: **M.Eng.** Carlos Javier Mojica Casallas

Universidad Santo Tomás

Facultad de Ingeniería Electrónica

Programa de Ingeniería Electrónica

Bogotá D.C

2019

**AGRADECIMIENTOS**

Esta tesis está dedicada 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**

**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

**Juan Manuel Silva**

##### **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.

Julian Ortiz

Contenido

[1. Resumen 9](#_Toc33807677)

[2. Planteamiento del problema 10](#_Toc33807678)

[3. Antecedentes 11](#_Toc33807679)

[4. Justificación 13](#_Toc33807680)

[5. Impacto Social 14](#_Toc33807681)

[6. Objetivos 15](#_Toc33807682)

[6.1. Objetivo general 15](#_Toc33807683)

[6.2. Objetivos específicos 15](#_Toc33807684)

[7. Marco teórico 16](#_Toc33807685)

[7.1. El aparato fonador 16](#_Toc33807686)

[7.2. Anatomía funcional de la voz 16](#_Toc33807695)

[7.3. Formantes 17](#_Toc33807705)

[7.4. Lóbulo parietal izquierdo 19](#_Toc33807706)

[7.5. Radionecrosis 20](#_Toc33807707)

[7.6. Introducción a la Teoría del Procesamiento Digital de Señales de Audio 20](#_Toc33807708)

[7.7. Transformada de Fourier 20](#_Toc33807709)

[7.8. Ventana de hamming 21](#_Toc33807710)

[7.9. Espectrograma 21](#_Toc33807711)

[7.10. Redes neuronales 22](#_Toc33807719)

[8. Diseño metodológico 24](#_Toc33807720)

[8.1. Diseño de prototipo 24](#_Toc33807723)

[8.2. Análisis y resultados 28](#_Toc33807724)

[8.2.1. Análisis de muestreo 28](#_Toc33807725)

[8.2.2. Procesamiento de audio 28](#_Toc33807726)

[8.2.2.1. Explicación de pre-procesamiento 28](#_Toc33807727)

[8.2.2.2. Explicación de método elegido para solucionar el proyecto 35](#_Toc33807728)

[8.2.2.3. Diseño de red neuronal de entrenamiento 35](#_Toc33807729)

[9. Análisis de resultados 38](#_Toc33807730)

[10. Conclusiones 43](#_Toc33807731)

[11. Bibliografía 44](#_Toc33807732)

Índice de ilustraciones

[Figure 1 Diagrama del aparato fonador 17](#_Toc34343447)

[Figure 2 Mediciones de los primeros formantes de las vocales del español. 18](#_Toc34343448)

[Figure 3 Determinación anatómica de las características de la voz 18](#_Toc34343449)

[Figure 4 Posición de la lengua al producir los fonemas de las vocales 19](#_Toc34343450)

[Figure 5 Lóbulo parietal izquierdo 20](#_Toc34343451)

[Figure 6 Ventana de hamming 21](#_Toc34343452)

[Figure 7 Espectrograma 22](#_Toc34343453)

[Figure 8 Diagrama de prototipo de captación de audio 24](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343454)

[Figure 9 Esquema de filtro implementado 25](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343455)

[Figure 10 Esquemático PCB 25](#_Toc34343456)

[Figure 11 Tarjeta de audio USB 26](#_Toc34343457)

[Figure 12 Tarjeta Raspberry Pi 3b+ 27](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343458)

[Figure 13 Parlantes 28](#_Toc34343459)

[Figure 14 Señal original 29](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343460)

[Figure 15 Señal reajustada con valores positivos 30](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343461)

[Figure 16 Delimitando señales de voz dentro de los audios 31](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343462)

[Figure 17 Recorte de voz de persona sana 32](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343463)

[Figure 18 Transformada de la señal graficada en la figura 11 33](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343464)

[Figure 19 Recorte de voz del paciente 33](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343465)

[Figure 20 Transformada de la señal graficada en la figura 12 34](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343466)

[Figure 21 Ventana de Hamming para señal producida por una persona sana 40](#_Toc34343467)

[Figure 22 Espectrograma de la señal producida por una persona sana 40](#_Toc34343468)

[Figure 23 Ventana de Hamming para señal producida por el paciente 41](#_Toc34343469)

[Figure 24 Espectrograma para señal producida por el paciente 42](#_Toc34343470)

[Figure 25 Diagrama de formantes producidas por vocal 42](#_Toc34343471)

Índice de tablas

[Tabla 1 Componentes utilizados en el prototipo 24](#_Toc34682886)

[Tabla 2 Cantidad de muestras por vocal 28](#_Toc34682887)

[Tabla 3 Pesos característicos del target para la red 35](#_Toc34682888)

[Tabla 4 Muestras entrenada por vocal 37](#_Toc34682889)

[Tabla 5 Pruebas de la red neuronal con las muestras del paciente 37](#_Toc34682890)

Índice de Código

[Código 1 Implementación en código de la ecuación 1 30](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343388)

[Código 2 Hallando las señales de voz dentro de la muestra de audio 31](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343389)

[Código 3 Librerías de la red neuronal 36](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343390)

[Código 4 Selección de modelo de red neuronal 36](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343391)

[Código 5 Optimizador de la red neuronal 37](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343392)

[Código 6 Modelo de almacenamiento de datos 37](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343393)

[Código 7 cargar la señal de entrada en el modelo de predicción 37](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343394)

[Código 8 Estimación de coeficientes LPC 39](file:///C:\Users\silva\OneDrive\Documentos\Monografia_ADELANTO.docx#_Toc34343395)

### 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 la gravedad 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 la transformada de Fourier y métodos como las redes neuronales y los filtros adaptativos.

### P**lanteamiento 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 **radionecrosis 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 **radionecrosis 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, y es producida por el aparato fonador que consta de tres partes:

* la cavidad infraglótica: en la cual se encuentran los órganos encargados de la respiración.
* la cavidad glótica (laríngea): donde se encuentra la laringe encargada de la fonación o el realizar el sonido.
* la cavidad supraglótica: en la cual se encuentra la cavidad bucal, la cavidad nasal y la faringe y es la encargada de resonar y articular el sonido.

Esta última es el área que tiene afectada el paciente lo que indica que las cuerdas bucales 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 al sensar las cuerdas bucales de una persona, mediante sensores situados al costado del cartílago cricoide.

### A**ntecedentes**

En el cuerpo humano la emisión del sonido consta de una vibración de las cuerdas bucales por impulso del aire, como se puede apreciar en el artículo [2].

El aparato fonador se divide en tres partes fundamentales:

* cavidad infraglótica: es el lugar donde se ubican los órganos respiratorios.
* cavidad glótica: principal órgano fonador, la laringe.
* cavidad supraglótica: se encuentran los órganos de la resonancia.

En el artículo del año 2017 de Turín Italia escrito por F. Casassa, A. Carullo y A. Vallan [3] se diseña un sistema que simula el aparato fonador mediante un dispositivo electrónico lo cual permite conocer el proceso de vibración de las cuerdas vocales y como se da la señal en el proceso de vibración, que normalmente no se puede apreciar lo que ofrece una ventaja para el desarrollo del proyecto al conocer la señal de salida en relación a la cantidad y velocidad de vibraciones.

El proyecto no plantea la falta de aparato fonador, por lo cual es ideal para conocer el comportamiento de las cuerdas bucales, debido a que la vibración de estas son las que producen la voz audible.

En el artículo del año 2009 de Lodz Polonia escrito por Damian krzesimowski y Zigmunt Ciota [4] se da a conocer la implementación de modelos matemáticos a una señal generada por una persona con dificultades de habla, por lo cual genera una señal diferente a la esperada, 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.

En este artículo se aprecian 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.

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 [5] 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 mediante un circuito electrónico que detecta el movimiento del glótico, además de realizar mediciones que me permiten mejorar la sensibilidad y calidad de la detección de la señal.

Este proyecto da una idea de cómo realizar mediciones que permitan mejorar la señal capturada por los sensores, debido a que al ser un sistema alámbrico presenta menor posibilidad de que haya pérdidas e interferencia con la señal censada.

En el artículo del año 2012 de Sivakasi India escrito por V. S. Balaji, N. R. Raajan y Har Narayan Upadhayay [6] se desarrolla 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 presente en el paciente permite conocer mejor el comportamiento de la señal sin realizar un procedimiento invasivo que genera más complicaciones al paciente con el objetivo de llegar posiblemente a la misma conclusión.

En el artículo del año 2017 publicado en NeuroImagen escrito por Martin Andermann, Roy D. Patterson, Carolin Vogt, Lisa Winterstetter y André Rupp [7] se explican las propiedades de la voz, como el ser inmune a la variación de forma de onda asociada. Además, explica el cómo se produce un sonido: la frecuencia del pulso global (GRP) que determina el tono vocal, y cuál fue el procedimiento a seguir para concluir lo anterior.

En el artículo del año 2018 publicado en Biomedical Signal Processing and control escrito por Susana Vaz-Freitas, Pedro Melo Pestana, Victor Almeida y Aníbal ferreira [8] mediante los parámetros acústicos más importantes de la señal que son: periódica, amplitud, duración y composición espectral. Estos parámetros caracterizan los atributos físicos de la voz en el tiempo.

Para lo anterior se realizaron pruebas bucales a 90 personas las cuales 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.

Este documento plantea la importancia de la frecuencia y la resolución en el procesamiento de la señal con el fin de obtener un buen resultado, además podemos encontrar parámetros acústicos como el Jitter para la comparación entre las señales del paciente y las de una persona sana.

EN este artículo del año 1972 publicado en Nokia Bell escrito por K. Ishizaka y J.L. Flagan [9] se realizó un modelo circuital equivalente de las glotis, permitiendo modelar matemáticamente su comportamiento, facilitando el entendimiento sobre la conducta de la señal generada al pasar el aire por la glotis permitiendo la vibración de las cuerdas lo que genera el sonido, también permite entender qué función cumple esta cavidad en temas de resonancia de la señal de la voz.

### Justificación

El cáncer es la segunda causa de muerte más frecuente a nivel mundial después de los ataques cardíacos [1] 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 **radionecrosis 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 una 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.

### 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 que el paciente pueda comunicarse de forma independiente nuevamente. La importancia del proyecto es que brinda 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 presentes diferentes problemas fonoaudiológicos, ya que esto permitiría que los pacientes mejoren su calidad de vida y logren comunicarse nuevamente

### Objetivos

### O**bjetivo 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.

### O**bjetivos 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 (radionecrosis 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 sensado 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.

### M**arco teórico**

Para comprender con mayor detalle este proyecto, es necesario aclarar algunos conceptos claves con el fin de poner en contexto sobre el problema real del paciente y las diferentes herramientas utilizadas para su solución.

### E**l aparato fonador**

En el artículo escrito por Herrera y Begoña (n.d.) [16] se encuentra una explicación más detallada de los órganos involucrados en la producción de sonido.

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.

El aparato fonador humano cumple con esos 3 elementos y se divide en tres partes fundamentales:

-la cavidad infla glótica: es el lugar donde se ubican los órganos respiratorios.

-la cavidad laríngea o glótica: se corresponde con el lugar ocupado por el principal órgano fonador, la laringe.

-la cavidad supra glótica: se encuentran los órganos de la articulación y la resonancia



### A**natomía funcional de la voz**

De acuerdo con el artículo de Dra. Begoña Torres (2007) [17]. La voz se produce mediante la acción conjunta de diferentes músculos como la laringe cuya parte esencial es la glotis, lo-s pulmones que se encargan 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.

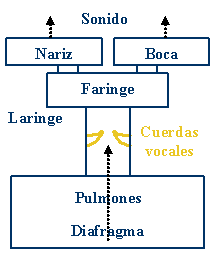


Figure 1 [Diagrama del aparato fonador](http://www.ehu.eus/acustica/espanol/musica/vohues/vohues.html)



### **Formantes**

Según el artículo [18] 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 #1) maneja un rango de baja frecuencia entre 200Hz 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.

Por otro lado, F2 (Formante #2) tiene un rango de operación superior a los 500 Hz hasta los 3500 Hz.

En resumen, F1 y F2 se convierten en bandas de frecuencias en donde se concentra la mayor parte de energía sonorante cuando se estimula las cuerdas vocales.

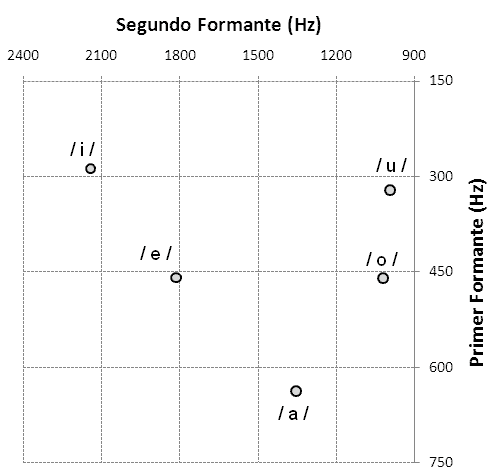


Figure 2 Mediciones de los primeros formantes de las vocales del español.

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, por ejemplo:

* En la vocal i se puede apreciar mayor distancia entre los valores de las frecuencias de los dos formantes.
* En las vocales e y o, presentan una frecuencia en F1 muy similar y en F2 se puede ver una gran variación entre ellas.
* Las vocales u y o, trabajan en un rango muy cercano en su frecuencia más alta (F2), pero varían mucho más en el Formante 1.

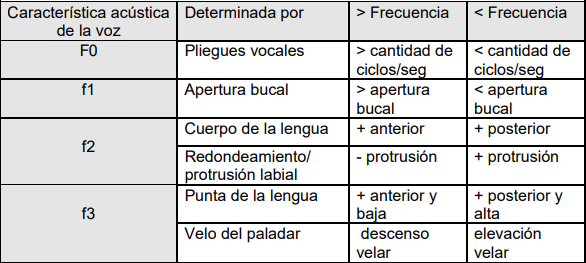


Figure 3 Determinación anatómica de las características de la voz

En la gráfica anterior podemos apreciar cuales partes del aparato fonador determinan las características, se puede observar que tanto el formante 1 esta relacionado a la apertura de la boca y el formante 2 a la forma de los labios encontrando el primer obstáculo debido a que el paciente presenta limitada movilidad en la zona de la boca debido a su diagnóstico.

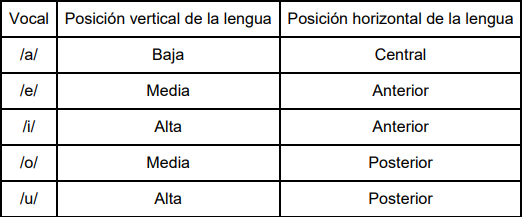


Figure 4 Posición de la lengua al producir los fonemas de las vocales

Adicionalmente para la producción de las vocales se debe tener una posición especifica de la lengua como se observa en la figura 4, la cual también es de vital importancia a la hora de 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.

### **Lóbulo parietal** **izquierdo**

Siguiendo el artículo [10] 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.

El lóbulo parietal izquierdo presenta más actividad en las personas diestras y este hemisferio es el encargado del procesamiento de la simbología de letras y números.

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.



Figure 5 Lóbulo parietal izquierdo

### **Radionecrosis**

Siguiendo el artículo [11] la radionecrosis se trata de una lesión cerebral producida por la radioterapia craneal en tumores o malformaciones vasculares cerebrales, esta consiste en la muerte de células.

Cuando la radionecrosis afecta el lóbulo parietal izquierdo puede dar lugar a lo que se llama “síndrome de gerstmann”. 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).

### I**ntroducción a la Teoría del Procesamiento Digital de Señales de Audio**

De acuerdo con el artículo [12] el procesamiento de señales es una de las tecnologías más poderosas, y se distingue de otras áreas de las ciencias computacionales por el tipo de datos que utiliza (señales). La representación de señales de audio analógicas como una secuencia de números es lograda por un conversor análogo-digital para así conseguir la generación de un sonido en una computadora ya que esta genera un modelo subyacente, debido a que, al transmitirse, mucha de la información se minimiza.

Con el análisis de phase vocoder es posible separar la información espectral de la información temporal.

### T**ransformada de Fourier**

Con lo visto en el artículo [13] la transformada de Fourier es una herramienta matemática útil para transformar señales entre el dominio del tiempo y el dominio de la frecuencia, con el fin de simplificar un problema en ecuaciones algebraicas, también es vista como un conjunto de proporciones acerca de fenómenos físicos.

Para que a una señal se le puede aplicar la transformada de Fourier esta debe cumplir con tres condiciones.

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

### Ventana de hamming

Según lo explicado en el artículo [19] las ventanas son funciones matemáticas utilizadas para el procesamiento de señales para evitar discontinuidades al principio y final de los bloques.

Gracias a esta función la ventana de hamming opera cancelando el lóbulo lateral más cercano, esto con el fin de hacer un análisis en una señal con una longitud limitada, en la figura 6 se observa un ejemplo de la función descrita.

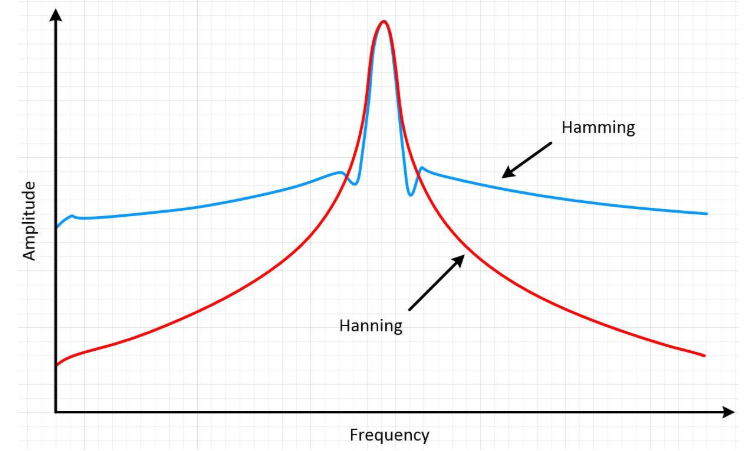


Figure 6 Ventana de hamming

### Espectrograma

Existen rangos de frecuencias sonoras que el oído humano no es capaz de percibir. Según lo visto en el artículo [19] gracias al espectrograma es posible ver las características que se presentan en frecuencias inferiores a 20Hz y superiores a 20KHz debido a que el espectrograma es una representación gráfica del espectro de frecuencias de emisión sonora.

Este tipo de graficas es mayormente utilizado para el estudio del comportamiento de la voz humana y registros bioacuáticos.

La forma en que se obtiene una de estas graficas como la observada en la figura 7 es mediante el uso de la trasformada de Fourier al principio de la señal a analizar, dependiendo del tamaño de la ventana utilizada para el análisis de la señal, se obtiene diferentes niveles de resolución. Si la ventana utilizada resulta ser muy pequeña resulta en una resolución baja ocasionando que no se pueda distinguir los armónicos dela señal.

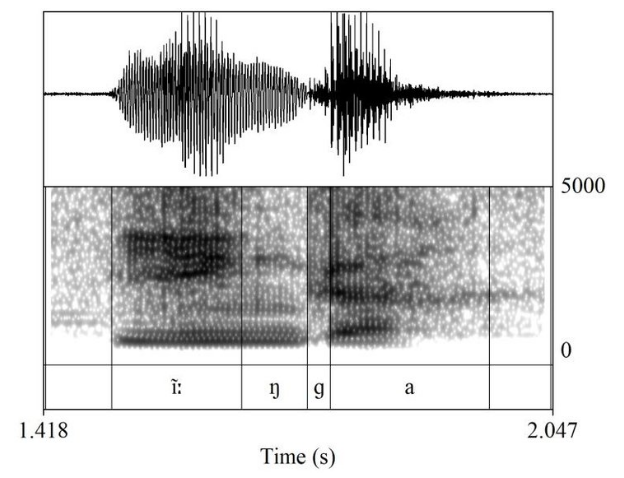


Figure 7 Espectrograma



### Redes neuronales

De acuerdo con el artículo [14] las redes neuronales son modelos matemáticos simplificados que simulan el procesamiento de un cerebro humano, que es normalmente utilizado para el reconocimiento de patrones de cualquier tipo. Estas unidades de procesamiento se organizan en 3 capas: la primera es la capa de 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. Por último, encontramos la capa 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 escoge que tipo de aprendizaje va a tener la red neuronal las cuales pueden ser supervisado, no supervisado y por esfuerzo.

* 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 de esta formar la red neuronal pueda encontrar los patrones presentes en la nueva información comparándolos con los datos aprendidos anteriormente.

### D**iseño metodológico**



### D**iseño de prototipo**

Para el diseño del prototipo fueron necesarios los siguientes materiales:

#### **Tabla 1 Componentes utilizados en el prototipo**

|  |  |
| --- | --- |
| **ELEMENTO** | **CANTIDAD** |
| **Raspberry pi3** | **1** |
| **Micrófono electret** | **2** |
| **Resistencia** | **2** |
| **Condensador** | **2** |
| **Parlante** | **2** |
| **Jack 3.5** | **2** |
| **Adaptador de audio usb** | **1** |
| **Conector AK500** | **3** |
| **Memoria micro sd 64 GB** | **1** |

En la siguiente imagen se presenta el esquema del prototipo.

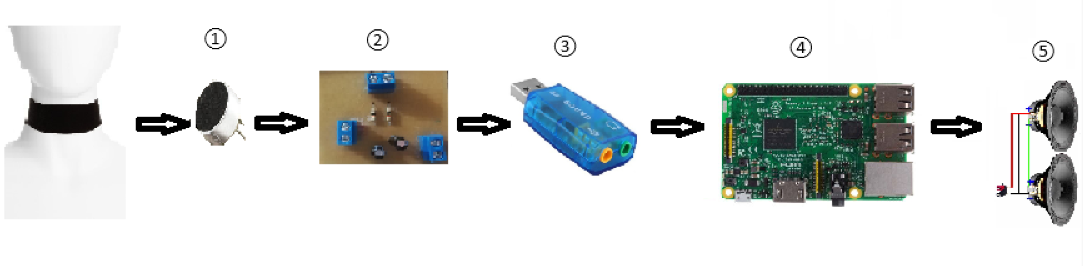


Figure 8 Diagrama de prototipo de captación de audio

El componente (1) es un micrófono electret con el que se llevó a cabo el censado de las vibraciones de las cuerdas vocales, este dispositivo tiene la ventaja de que por su tamaño permite la captación de una forma menos invasiva sobre el paciente, para el funcionamiento de este es necesario incorporar un pequeño circuito con una alimentación entre 2v - 9v.

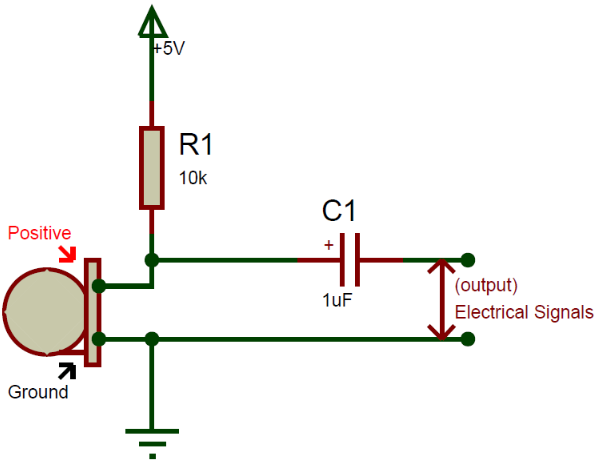
El componente (2) es un filtro que está compuesto de dos resistencias de 10kᘯ esto con el fin de limpiar la corriente que fluye a través del micrófono y dos condensadores de 1uF que filtran el ruido de CC (corriente continua), como se muestra en la figura 4.

Figure 9 [Esquema de filtro implementado](https://components101.com/electret-condenser-microphone)

Imagen que contiene texto

Descripción generada automáticamente

Figure 10 Esquemático PCB

En la figura 10encontramos una descripción un poco más detallada del esquema utilizado para el de la PCB en la alimentación de los micrófonos.

El componente (3) que se observa en la figura 5 es un adaptador de audio USB, este dispositivo permite leer y almacenar archivos de audio, gracias a este proceso el microprocesador puede dedicar más recursos en actividades más complejas.



Figure 11 Tarjeta de audio USB

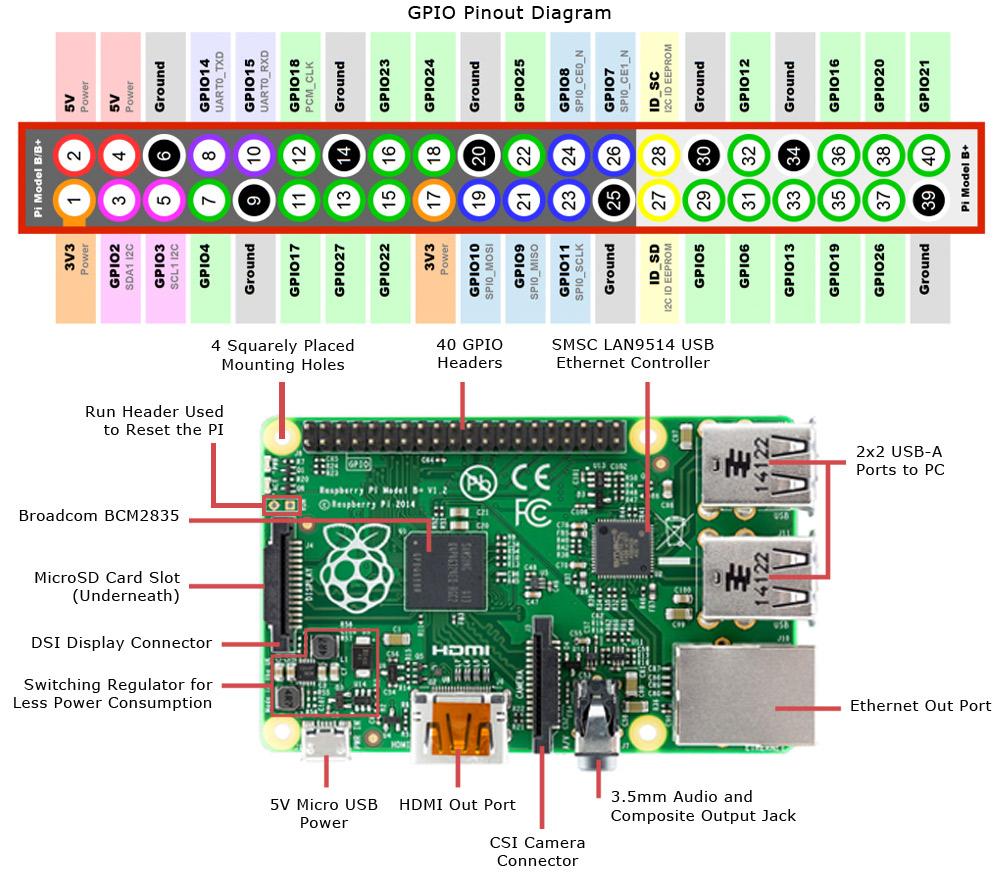
El componente (4) es una tarjeta con un sistema embebido también conocida como raspberry pi3, este dispositivo fue utilizado por su capacidad de procesamiento de información además de su posibilidad de aumentar el almacenamiento del dispositivo en caso de ser necesario, en este prototipo se utilizó una memoria micro sd de 64 GB, esta tarjeta está compuesta por diferentes módulos como se muestra en la siguiente figura.

Figure 12 [Tarjeta Raspberry Pi 3b+](https://3.bp.blogspot.com/-xYpPeD5bht4/WDGhmkeyaNI/AAAAAAAABGM/sCLEer-ndJwhG8OJS2s3CaOtIsY2WQMBACLcB/s1600/pi_board_pinout.jpg)

Los módulos que se utilizados fueron los de USB con estos se integró el componente (3) para realizar el debido procesamiento de los datos obtenidos.

por último, el componente (5) que se ve en la figura7 son los parlantes que reproducen la voz artificial generada por el componente (4).



Figure 13 [Parlantes](https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwjHiYn9q_fkAhVqhuAKHY74BN8QjRx6BAgBEAQ&url=https%3A%2F%2Flistado.mercadolibre.com.ar%2Fparlante-0.5w&psig=AOvVaw1Q1Exld3lv2XBPASnfwFYz&ust=1569891564938622)

### A**nálisis y resultados**

### A**nálisis de muestreo**

Para la selección del dispositivo encargado de sensar las cuerdas vocales, se tomó en cuenta la comodidad del paciente de manera que fuera lo menos invasivo posible. Al momento de realizar los muestreos, se evidencia que el paciente toma como guía los gestos efectuados al momento de pronunciar una vocal, debido a que intenta reproducir la gesticulación y, por consiguiente, mejorar la calidad de la señal producida. Sin embargo, debido a la dificultad que presenta el movimiento de los músculos de la boca que intervienen en el proceso de fonación, se produce una gran variación en los sonidos, lo cual requiere una base de datos muy grande que permita tener la mayor cantidad de muestras posibles y de esta manera disminuir el porcentaje de error que pueda generar dicha variedad en el sistema.

La cantidad de muestras que se obtuvieron para realizar el procesamiento se aprecian

#### **Tabla 2 Cantidad de muestras por vocal**

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

### Procesamiento de audio

### E**xplicación de pre-procesamiento**

Este procedimiento se realiza con el fin de analizar mediante el software Matlab las diferencias que presenta la señal producida por una persona sana a la producida por el paciente, mediante el análisis de gráficas en tiempo y frecuencia para determinar los posibles métodos de solución al problema planteado.

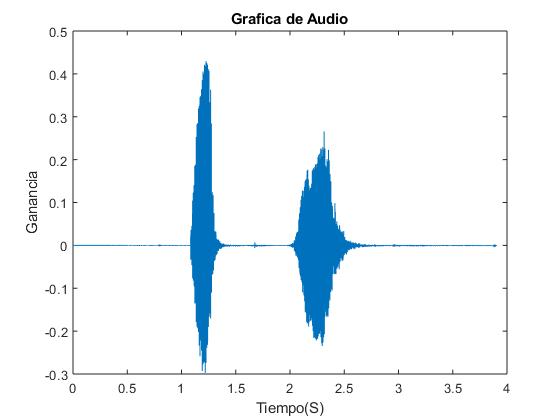
Se procede a leer un archivo de audio en el cual se encuentra la voz de una persona sana seguida por la voz del paciente, permitiendo conocer datos como la frecuencia de muestreo, la duración del audio, así como las ganancias producidas por las voces, gracias a lo cual se realiza una gráfica de ganancia contra tiempo como se aprecia en la figura 14.

Figure 14 Señal original

A continuación, se normaliza la señal y se procede a trabajar con los valores de ganancia positivos, aplicando la siguiente fórmula que se ve en la ecuación 1:

Ecuación 1

se implementa en Matlab de la siguiente manera que se muestra en el código 1

for i=1:N-Nf

y2(i)=sum(y1(i:i+Nf-1))/500;

end

Código 1 Implementación en código de la ecuación 1

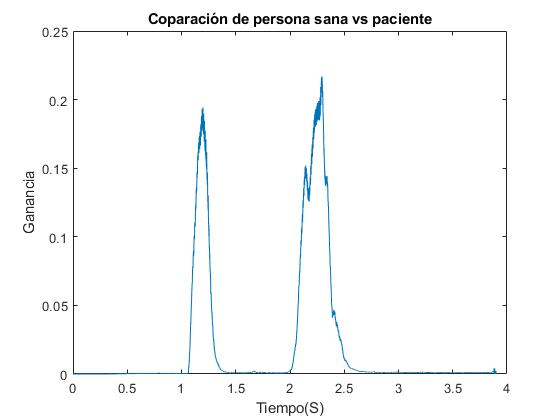
Siendo y1 el arreglo con la información de la señal, n el tamaño de la muestra de audio, obteniendo como resultado y2 que al ser graficada en valor absoluto para solo observar la parte positiva de la señal da como resultado la imagen de la figura 9

Figure 15 Señal reajustada con valores positivos

Con la información de la figura 15 se puede despreciar el ruido del medio debido a que las ganancias que presentan son bajas comparadas con las ganancias de las voces en la señal, gracias a lo cual se observa en que partes de la señal esta la información relevante como se ve en la figura16

Para realizar la extracción de la señal del paciente de las muestras de audio procedemos a encontrar el valor de ganancia presente en el paciente, pero no en la persona sana de la siguiente manera en Matlab

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)))

plot(yJ)

ciclos=floor(length(yJ)/2048)

if(ciclos==0)

aux=aux+0.01

end

end

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

* y2 el arreglo de la señal después de procesarlo con la ecuación 1
* 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
* aux es un auxiliar que permite comparar las ganancias
* Ciclos determina si la cantidad de datos presentes en yJ son suficientes para realizar la transformada de la señal respectiva.

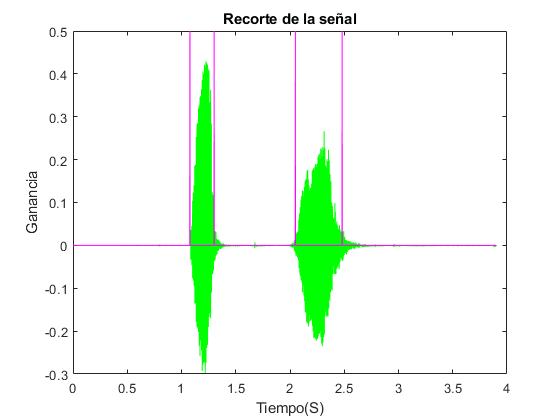


Figure 16 Delimitando señales de voz dentro de los audios

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.

En las siguientes graficas podemos apreciar la diferencia de las señales obtenidas en el código 2 del paciente y de una persona sana, así como los valores de ganancia producidos por estos

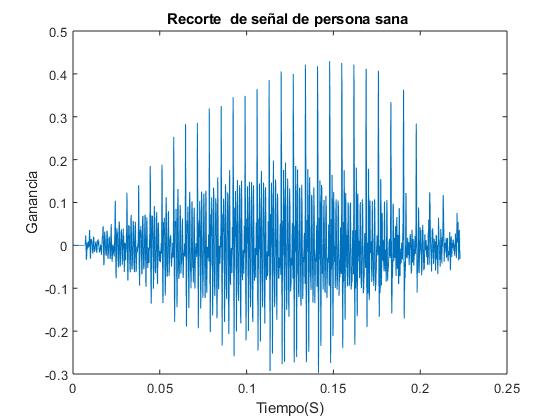


Figure 17 Recorte de voz de persona sana

Podemos ver como la señal producida por la persona sana se puede ver como 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.

A las señales obtenidas se le realiza una transformada de Fourier que pasa del dominio del tiempo al dominio de la frecuencia tomando solo la mitad de la señal producida debido a que se refleja al final obteniendo la gráfica de la figura 19

A las señales obtenidas se le realiza una transformada de Fourier que pasa del dominio del tiempo al dominio de la frecuencia tomando solo la mitad de la señal producida debido a que se refleja al final obteniendo la gráfica de la figura 19

A las señales obtenidas se le realiza una transformada de Fourier que pasa del dominio del tiempo al dominio de la frecuencia tomando solo la mitad de la señal producida debido a que se refleja al final obteniendo la gráfica de la figura 19

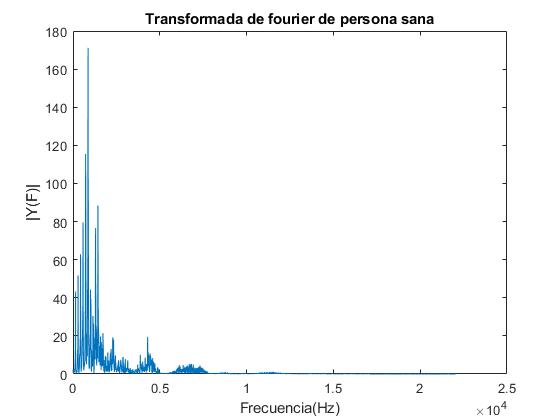


Figure 18 Transformada de la señal graficada en la figura 11

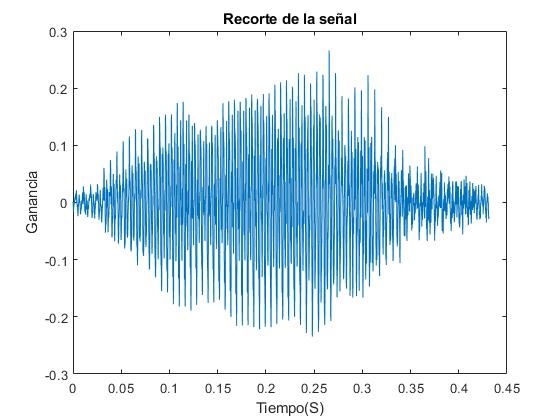


Figure 19 Recorte de voz del paciente

Entre las características que se pueden diferenciar son las variaciones de los picos de los valores de |y(f)|, donde se aprecia que las imposibilidades de mover los músculos de la boca a voluntad inducen en que este valor aumente en 166% aproximadamente, lo cual podría producir errores a tener en cuenta en la caracterización de la señal.

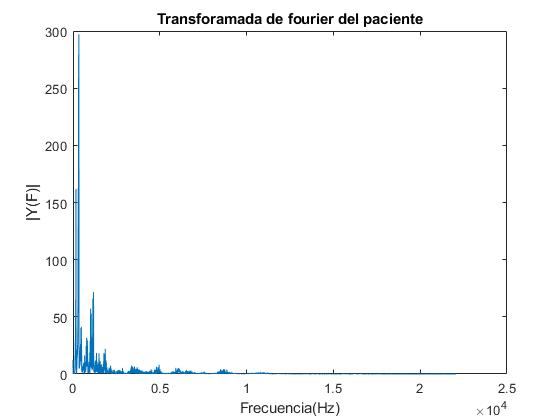


Figure 20 Transformada de la señal graficada en la figura 12

ara observar mejor la señal se procede a realiza la gráfica de la transformada en 3d en la figura (pendiente número de figura) podemos observar los picos más altos de frecuencia se encuentra la mayoría de la información.

En la figura se puede observar la transformada de Fourier en 3d donde se aprecian los picos de ganancia más fuertes

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

En el dominio de la frecuencia, la detección de la frecuencia fundamental trata de encontrar la presencia de frecuencias con mayor intensidad para encontrar así los armónicos más cercanos a la frecuencia fundamental.

### E**xplicación de método elegido para solucionar el proyecto**

se tiene una base de datos que contiene 376 muestras las cuales son medidas y procesadas para crear un archivo (.cvs), el proceso elegido fue una red neuronal debido a su capacidad de procesamiento y de reconocimiento de patrones.

Para este procesamiento se implementó una red de clasificación que tiene como objetivo analizar la señal obtenida del paciente mediante el censado y procesarlo para encontrar patrones que le permita relacionar la señal producida con algunas de las presentes en la base de datos

Se seleccionan redes neuronales debido a que son algoritmos matemáticos que permiten procesar datos de una manera más versátil; ya que, facilitan la búsqueda de patrones. Esto es muy útil en este caso, puesto que la enfermedad del paciente dificulta la fonación y no se presentan similitudes evidentes entre las señales producidas por una misma vocal; por lo que, el uso de redes neuronales permitirá identificar ciertas frecuencias en las que se supone se encuentra la información de las vocales pronunciadas por el paciente. Además, se adicionan 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 frecuencias, que permite el paso de una mayor cantidad de información útil. Para calcular dichas frecuencias, es necesario tener en cuenta los coeficientes arrojados por las redes neuronales; dichos coeficientes son resultado del entrenamiento de la red neuronal y dependen de los valores de entrada de la señal.

### D**iseño de red neuronal de entrenamiento**

Para armar el datase de la red neuronal se procede a almacenar en un archivo con formato CVS, una matriz en la cual se almacena la muestra y el número de la muestra y la señal del paciente pre-procesado, además en un archivo .dat se almacena un vector que presenta un coeficiente para cada vocal como se observa en la Tabla 3

#### **Tabla 3 Pesos característicos del target para la red**

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

A continuación se carga los archivos .cvs y .dat, para ser asignado a una variable cada uno luego se extraen los modelos de redes neuronales de la librería keras.

from sklearn.model\_selection import train\_test\_split

from keras.utils import to\_categorical

Código 3 Librerías de la red neuronal

Para escoger el tipo de red que vamos a entrenar tenemos los siguientes comandos

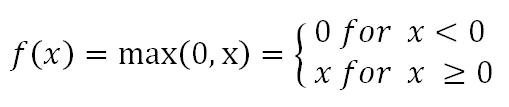
Código 4 Selección de modelo de red neuronal

classifier.add(Dense(output\_dim=11, init='uniform',activation='relu',input\_dim=50000 ))

classifier.add(Dense(output\_dim=9, init='uniform',activation='relu' ))

La cual es una red clasificadora teniendo en cuenta el comportamiento de las señales obtenidas, donde la función de activación seleccionada es una función relu (unidad rectificadora lineal) que se representa con la siguiente función matemática presentada en la ecuación 2

Ecuación 2



Permitiéndome despreciar los valores negativos de forma que el aprendizaje de la red solo tome en cuenta los coeficientes positivos mejorando el reconocimiento.

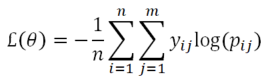
Código 5 Optimizador de la red neuronal

classifier.compile('adam', loss='categorical\_crossentropy', metrics=['accuracy'] )

El optimizador de la red diseñada es Adam debido a que adapta el radio de aprendizaje de acuerdo a la distribución de los parámetros de aprendizaje, por lo que entre más dispersos se encuentren los datos mayores es el radio de aprendizaje.

Además, para la estimación de error de la red se seleccionó el estimador Categorical Cross-Entropy debido a su precisión para variables categóricas, su representación matemática se observa en la ecuación 3

Ecuación 3



Por ultimo en el código 6 se puede apreciar que en la métrica se seleccionó accueracy que es el encargado de averiguar la cantidad de aciertos donde el modelo predice el resultado de forma correcta.

from keras.models import model\_from\_json

Código 6 Modelo de almacenamiento de datos

Se carga el modelo json dese la librería de keras con la cual se puede crear una pequeña biblioteca de datos.

En el programa principal se leer el modelo de predicción guardado en el archivo json, acto seguido se capta la señal de sonido con el prototipo a la cual se le aplica el pre procesamiento, la señal resulte se ingresa al modelo de predicción como se observa en el código 7

entrada[0,0:tamm1]=Yj[0:tamm1]

result = loaded\_model.predict(entrada)

result=np.array(result)

print(result)

Código 7 cargar la señal de entrada en el modelo de predicción

El modelo de predicción da como resultado unos pesos dentro de los cuales el valor más grande es la predicción que la red establecido para la señal de audio cargada en el modelo.

#### **Tabla 4 Muestras entrenada por vocal**

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

### Análisis de resultados

De la red entrenada se obtuvieron los siguientes resultados

#### **Tabla 5 Pruebas de la red neuronal con las muestras del paciente**

|  |  |  |
| --- | --- | --- |
| Vocal | Número de pruebas | Porcentaje de aciertos |
| A | 150 | 40% |
| E | 140 | 32% |
| I | 75 | 31% |
| O | 125 | 35% |
| U | 130 | 33% |

Para realizar el análisis de los resultados obtenidos por la red se procede a realizar un análisis de análisis espectral de las señales, el cual nos dé más claridad sobre como el diagnóstico del paciente afecta la producción de los sonidos de los fonemas.

Tomando la misma muestra de la señal expuesta en el pre-procesamiento para realizar la comparación de las formantes presentes en la señal.

Para poder realizar este procedimiento se debe realizar mediante LPC (linear predictive code), siendo un análisis de series temporales conocido como modelo Autorregresivo de la siguiente forma observada en la ecuación 4

Ecuación 4

Donde Sn es una serie temporal dada como una combinación lineal de sus p valores previos y de una entrada Un

Al cual se le aplica una transformada de Fourier para obtener

Ecuación 5

Done ak es el coeficiente de predicción lineal y la G una ganancia

Calculo del coeficiente de auto correlación se evidencia en la ecuación 6

Ecuación 6

La implementación de las ecuaciones anteriormente desarrolladas se presenta a continuación

A = lpc(xJ1,16);

rts = roots(A);

rts = rts(imag(rts)>=0);

Código 8 Estimación de coeficientes LPC

Haciendo uso de los módulos de Matlab se implementan estos cálculos para obtener la ventana de Hamming de las Figuras 21 y 23, el las cuales se puede observar la ventana de Hamming de color rojo, la señal inicial de color verde, la señal azul representa la aplicación de la ventana a la señal y por último el color morado es la señal después de aplicarle un filtro pasa altos



Figure 21 Ventana de Hamming para señal producida por una persona sana

El espectrograma nos permite hacer un análisis visual de los valores en los que se encuentran las frecuencias fundamentales



Figure 22 Espectrograma de la señal producida por una persona sana

El espectrograma de la señal producida por una persona sana apreciado en la figura 22 nos muestra que nos valores de ganancia sobre frecuencia se encuentran entre los 50 a 70 DB/Hz y ubicado entre 1 y 1.4 segundos.

En la figura 23 se observa como el filtro pasa altos, desprecia varios valores de ganancia de la señal producida por el paciente, lo cual puede deberse a la resonancia acústica producida por la disminución de movimiento de músculos como los labios o la lengua



Figure 23 Ventana de Hamming para señal producida por 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



Figure 24 Espectrograma para señal producida por el paciente



Figure 25 Diagrama de formante

s producidas por persona sana para la vocal E

La figura 25 presenta la gráfica de los formantes 1 y 2 los cuales son las más importantes para diferenciar las vocales producidas por una persona, en el eje x se encuentra el formante 1 y en el eje y se encuentra el formante 2.

Las personas cuando hablan producen las formantes siempre en un rango por lo cual son ideales para realizar la caracterización de las vocales. Debido a esto se realizó la grabación de varias muestras de audio de la misma persona para conocer la variación en el formante de la voz y tener una idea de los mayores rangos producidos, después de lo cual se procede calcular el promedio de todos los muestreos de la misma vocal para conocer.



Figure 26 Diagrama de formantes producidas por el paciente para la vocal E

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.

En la figura 26 podemos ver como la disminución en los movimientos de la cavidad vocal aumenta el valor de la frecuencia fundamental 1 alejándola del valor promedio calculado en el que se ubica la vocal o para un hombre.

### C**onclusiones**

* El desarrollo del dispositivo de captación de señales de audio fue diseñado para no ser intrusivo, lo que nos permite utilizar tarjetas de procesamiento robustas, así como tener un tamaño mucho mayor lo que permite diseñar más etapas de pre procesamiento como filtros que ayuden a disminuir la carga del pre procesamiento realizado por el software.
* La eficiencia con la que la red neuronal trabaje recae en la constancia y claridad con el que el paciente repita los patrones en los que emite una vocal, esto debido a que la red esta entrenada en encontrar ciertos patrones recurrentes que le permitan identificar la vocal producida, por lo cual al paciente tener el control total de los músculos que intervienen en el proceso de fonación es muy difícil modificar y controlar de forma consiente las señales producidas por el paciente.
* El proceso reiterado de habla y el sobre esfuerzo a la hora de producir sonidos produce una fatiga muscular que incrementa la afecta de la señal producida por el paciente, lo cual tiene una incidencia en la ganancia y frecuencia a la que se produce la señal, lo que a su vez dificulta la captación y el procesamiento de las muestras obtenidas.
* Los resultados obtenidos a lo largo del trabajo demuestran la complejidad de los órganos humanos y su funcionamiento, y la limitada cantidad de conocimiento que poseemos sobre fenómenos como el habla, y cono el estudio de áreas como la acústica y la electrónica nos permiten realizar nuevos descubrimientos acerca de los procesos que nos permiten hablar en el dia a dia

### **Bibliografía**

|  |  |
| --- | --- |
| [1] | 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]. |
| [2] | B. M. M. Cristian Herrera Hernandez, «virtuniversidad,» [En línea]. Available: <https://www.virtuniversidad.com/greenstone/collect/ingles/index/assoc/HASHf01b/45450734.dir/doc.pdf>. [Último acceso: 12 abril 2018]. |
| [3] | 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. |
| [4] | Damian Krzesimowski, «Voice Signal Processing for Patiens with Stroke Hospitalisation,» de Mixed Design of Integrated Circuits & Systems, 2009. MIXDES '09. MIXDES-16th International Conference, Lodz, Polonia, 2009. |
| [5] | 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. |
| [6] | N. R. R. H. N. U. V. S. Balaji, «Identification of Predominent 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. |
| [7] | 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. |
| [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] | 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. |
| [10] | «psicoactiva,» [En línea]. Available: https://www.psicoactiva.com/blog/lobulo-parietal-anatomia-funcion/. [Último acceso: 2019 09 29]. |
| [11] | N. Contemporánea, «Neurocirugía Contemporánea,» 2008. [En línea]. Available: http://neurocirugiacontemporanea.com/doku.php?id=radionecrosis. [Último acceso: 29 09 2019]. |
| [12] | «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]. |
| [13] | L. d. a. d. f. aplicados, «transformada de fourier,» [En línea]. Available: http://www4.ujaen.es/~jmalmira/transformada\_fourier\_almira.pdf. [Último acceso: 2019 09 2019]. |
| [14] | 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]. |
| [15] | Ecured, «Algoritmo de clasificación,» [En línea]. Available: https://www.ecured.cu/Algoritmos\_de\_clasificaci%C3%B3n\_supervisada. [Último acceso: 29 09 2019] |

|  |  |
| --- | --- |
| [16] | Herrera, C., & Begoña, M. (n.d.), «El aparato fonador. Física- Fisiología II.,» [En línea]. Available: <https://www.virtuniversidad.com/greenstone/collect/ingles/index/assoc/HASHf01b/45450734.dir/doc.pdf>. [Último acceso: 29 09 2019] |
| [17] | Dra. Begoña Torres, «Anatomía funcional de la voz» [En línea]. Available: <http://www.medicinadelcant.com/cast/1.pdf>. [Último acceso: 29 09 2019] |
| [18] | musiki, «musiki» [En línea]. Available: http://musiki.org.ar/Formantes. [Último acceso: 23 09 2019] |
| [19] | SCRIBD, «SCRIBD» [En línea]. Available: <https://es.scribd.com/document/358481883/Ventanas-de-Hamming> [Último acceso: 23 09 2019] |
| [20] | musiki, «musiki» [En línea]. Available: <http://musiki.org.ar/Espectrograma_(sonograma)> [Último acceso: 25 07 2019] |
|  |  |
| [21] | musiki, «musiki» [En línea]. Available: <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> [Último acceso: 25 07 2019] |
| [21] | musiki, «musiki» [En línea]. Available:  <http://iie.fing.edu.uy/ense/asign/tes/monografias/2010/iturralde-rosa/estimacion_y_seguimiento-iturralde_rosa.pdf>  [Último acceso: 25 07 2019] |