



# **Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.**

**Ivie Dayana Yépez Vargas**

Universidad Antonio Nariño  
Facultad de Ingeniería Mecánica, Electrónica y Biomédica  
Pasto, Colombia  
2019



# **Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.**

**Ivie Dayana Yépez Vargas**

Proyecto de grado monográfico presentado como requisito parcial para optar al título de:  
Tecnólogo en mantenimiento en electromecánica

Director (a):

Ingeniero Luis Enrique Arteaga

Línea de Investigación:

Automatización

Universidad Antonio Nariño

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Pasto, Colombia

2019



## **Agradecimientos**

En primer lugar agradecer a Dios, el poder infinito que nos ayuda a lograr nuestras metas, y quien nos alienta con coraje para no desfallecer con el fin de superarnos a nosotros mismos día a día, agradecer a nuestra familia por el apoyo y el ánimo que me inculcaron de perseverar, agradezco a nuestros profesores por los conocimientos transmitidos, basados en su ejemplar aptitud de enseñanza, a nuestros compañeros por compartir los difíciles y exitosos momentos, gracias a todas las personas que han aportado desde el más mínimo detalle, y que han caminado a nuestro lado invirtiendo su tiempo y dedicación absoluta para el alcance nuestros logros.



## Resumen

Este proyecto se basa en la investigación sobre sistemas de reconocimiento de comandos de voz, con el fin de crear un algoritmo que permita optimizar el libre desarrollo de personas con alguna discapacidad motriz, utilizando su voz para controlar una silla de ruedas. Se exploró una notación capaz de detectar 12 características fundamentales del espectro, empleando los modelos ocultos de Markov, cuyo propósito era obtener datos tras la observación de eventos, funciones probabilísticas, caracterizadas por un conjunto finito de  $N$  estados en la cadena de primer orden, y el conjunto de probabilidades de transición entre etapas, denotando los instantes de tiempo asociados a los cambios de estado, siendo capaz de procesar las características y factores necesarios con el fin de llevar a cabo el esbozo investigativo de la presente monografía.

**Palabras clave:** comandos de voz, reconocimiento, procesamiento de señales.

## Abstract

In this project an investigation is carried out on systems for the recognition of voice commands, which is adapted to a wheelchair, this approach is carried out taking into account the population with motor disability, which is affected by the Exclusive society in terms of providing services, spaces, work, etc. The purpose of our work lies in the development of an algorithm that summarizes the optimal methodologies, extracting the necessary techniques for a good processing of acoustic signals, capable of filtering inaudible and harmful frequency signals for the human being, where the range of usefulness is from 20Hz to 20kHz, this algorithm is capable of recognizing no less than 12 fundamental characteristics of the voice spectrum, in order to identify the phonemes of the words. Among the methods that are part of the system is the Hidden Markov Models, which are statistical comparisons.

**Keywords:** voice commands, recognition, signal processing



# Contenido

Pág.

Resumen.....	VII
Abstract.....	VIII
<b>Contenido</b> .....	<b>IX</b>
Lista de figuras.....	11
Lista de Símbolos y abreviaturas.....	12
Introducción.....	14
1. Objetivos .....	16
1.1    Objetivo general .....	16
1.2    Objetivos específicos .....	16
2. Metodología.....	17
3. Estado del arte .....	18
3. 1 Discapacidad motriz y Tecnología.....	18
4. Marco teórico.....	20
4.1 Fisiología de la voz .....	20
4.2 Clasificación de los Sonidos.....	22
4.3 Fisiología de la Fonación .....	23
4.4 Oído humano .....	25
4.5    Sistemas para la adquisición de señales.....	27
4.5.1 Características de las ondas de sonido .....	28
4.6 Estructura para el reconocimiento de comandos de voz .....	28
4.6.1 Pre-procesamiento.....	28
4.6.2 Pre énfasis .....	29
4.6.3 Segmentación .....	30
4.6.4 Enventanado .....	31
4.6.5 Extracción de características .....	31

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

• Banco de filtros de Mel .....	35
• Transformada de cosenos discreta de Fourier .....	36
4.6.6. Reconocimiento de voz .....	36
• DTW Dinámico Time Warping .....	37
• Modelos ocultos de Markov (HMMs) .....	37
• Reconocimiento de voz registro de sonidos .....	38
• Reconocimiento del habla basado en fonemas .....	38
5.1.1 Pre procesamiento .....	40
• Preénfasis .....	40
• Segmentación .....	40
• Enventanado .....	40
5.1.2 Extracción de características .....	41
5.1.3 Reconocimiento de voz .....	41
5.2 Elaboración de un algoritmo para reconocimiento de comandos de voz .....	42
5.2.1 Redacción .....	42
5.3 Diagrama de flujo .....	44
6. Conclusiones y recomendaciones .....	45
6.1 Conclusiones .....	45
6.2 Recomendaciones .....	46
Bibliografía .....	47

## Lista de figuras

Pág.

<b>Figura 1:</b> Representación de ondas.....	21
<b>Figura 2:</b> Aparato Fonador.....	21
<b>Figura 3:</b> Esquema Aparato Fonador.....	22
<b>Figura 4:</b> Niveles de Apertura Glotal .....	23
<b>Figura 5:</b> Puntos de Articulación .....	24
<b>Figura 6:</b> Modelo de Producción de Voz .....	25
<b>Figura 7:</b> Anatomía del oído humano .....	26
<b>Figura 8:</b> Estructura Interna del Pre-Procesamiento.....	29
<b>Figura 9:</b> Descomposición de una señal compleja en sumatoria de señales simples.....	32
<b>Figura 10:</b> Transformada de Fourier.....	34
<b>Figura 11:</b> Diagrama FFT para N=8.....	35
<b>Figura 12:</b> Estructura interna del procesamiento.....	36
<b>Figura 13:</b> Estructura de un sistema de reconocimiento de voz.....	39
<b>Figura 14:</b> Estructura final de un sistema de reconocimiento de voz.....	40
<b>Figura 15:</b> Extracción de Características.....	41
<b>Figura 16:</b> Estructura de reconocimiento de comandos de voz.....	42
<b>Figura 17:</b> Diagrama de flujo.....	44

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

## Lista de Símbolos y abreviaturas

A continuación, se presentan algunos ejemplos.

### Símbolos con letras latinas

Símbolo	Término	Unidad SI	Definición
$A$	Área	$m^2$	$\iint dx dy$
$A_{BET}$	Área interna del sólido	$\frac{m^2}{g}$	ver DIN ISO 9277
$A_g$	Área transversal de la fase gaseosa	$m^2$	Ec. 3.2
$A_s$	Área transversal de la carga a granel	$m^2$	Ec. 3.6
$a$	Coefficiente	1	Tabla 3-1

### Símbolos con letras griegas

Símbolo	Término	Unidad SI	Definición
$\alpha_{BET}$	Factor de superficie	$\frac{m^2}{g}$	$(W_{F,waf})(A_{BET})$
$\beta_i$	Grado de formación del componente i	1	$\frac{m_j}{m_{bm} \varphi}$
$\gamma$	Wandhaufreiwinkel (Stahlblech)	1	Sección 3.2
$\varepsilon$	Porosidad de la partícula	1	$1 - \frac{\rho_s}{\rho_w}$
$\eta$	mittlere Bettneigungswinkel (Stürzen)	1	Figura 3-1

### Subíndices

Subíndice	Término
bm	Materia orgánica
DR	Dubinin-Radushkevich
E	Experimental

### Superíndices

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

**Superíndice Término**

---

N Exponente, potencia

**Abreviaturas****Abreviatura Término**

---

1.LT Primera ley de la termodinámica

DF Dimension fundamental

## Introducción

El presente estudio busca dar continuidad al trabajo de monografía realizado en conjuntamente señor WILLIAM RENE INGUILAN CEBALLOS en el año 2018, durante el Diplomado en Inteligencia Artificial y Procesamiento de Señales, realizado por el programa de Tecnología en Electromecánica de la Universidad Antonio Nariño; donde se inició este proceso investigativo, extrayendo características de participación binomial, que conllevo a sucesiones sistemáticas en el procesamiento de comandos de voz, elaboración algorítmica y resultados.

Teniendo en cuenta la distribución mundial de la población con discapacidades en relación con los patrones geográficos del desarrollo humano (*Cutillas 2016*), las personas con discapacidad se regulan según su escala geográfica afectando la organización social y territorial de las comunidades, causando afectación a personas con discapacidades y no todas estas tienen los recursos o las capacidades para resistir la exclusión y opresión social, (Abberley, 1987; Imrie, 1996; *Cutillas 2016*), de igual forma las personas con algún tipo de discapacidad motora, afronta diversos impedimentos y obstáculos en su libre desarrollo, debiéndose garantizar todos los servicios de salud e inclusión.

Es así que como la Ley Estatutaria 1608 de 2013 *“por medio de la cual se establecen las disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad”* genera en la Republica de Colombia medidas y acciones afirmativas que permiten a las personas con discapacidad, bajo un marco de corresponsabilidad, ejercer sus derechos en igualdad de condiciones con las demás personas.

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

Por ende, la Tecnología debe estar inmersa y hacer parte de este avance social que busca la igualdad en todos los ciudadanos por medio de la creación y desarrollo de equipos tecnológicos que lleven a la población tradicionalmente discriminada por sus condiciones motoras, a realizar acciones autónomas por medio de la implementación algorítmica del procesamiento de comandos de voz.

Los argumentos tratados en la presente monografía se basan en el desarrollo progresivo de obtención de señales de voz digitalizadas previamente, a las cuales se les organiza mediante parámetros lingüísticos, lo cual permite disminuir el ruido emitido por la misma fonética o de sus alrededores, logrando que su intensidad se atenúe, alcance y posibilite la extracción de características que coincidan en la voz humana, utilizando los resultados finales para el control de comandos de voz en sillas de ruedas, todo este concepto descriptivo es realizable gracias a los modelos ocultos de Marcov.

El presente trabajo recopila información teórica obtenida de la revisión bibliográfica de diferentes estudios realizados en el campo del procesamiento de voz, buscando aportar a futuras implementaciones realizadas por estudiantes universitarios, creadores científicos, fabricantes biomédicos y empresarios de baja y gran escala; que se interesen en desarrollar una silla de ruedas capaz de incluir de forma autónoma a personas con discapacidad motriz.

# **1. Objetivos**

## **1.1 Objetivo general**

Abordar el estudio del procesamiento de comandos de voz, empleando los modelos ocultos de Markov, con el fin de ser aplicado al control autónomo de una silla de ruedas electrónica.

## **1.2 Objetivos específicos**

Elaborar una revisión bibliográfica con contenidos basados en técnicas y estructura para el procesamiento de comandos de voz.

Analizar diferentes técnicas para determinar las más acertadas en el procesamiento de señales acústicas para la extracción de características de voz teniendo como principal referente los modelos ocultos de Markov.

Construir un algoritmo para el reconocimiento de comandos de voz aplicado al manejo de sillas de ruedas mediante la utilización de los modelos ocultos de Markov.



## **2. Metodología**

Por tratarse de una investigación monográfica, se considera de tipo documental exploratoria; donde se aborda la información referente al procesamiento de comandos de voz; enfatizando en el proceso de adquisición de señales, las características de las ondas de sonido y la estructura del sistema de reconocimiento de voz. Para finalizar, se presenta un algoritmo para el reconocimiento de comandos de voz. En este caso, se adoptó el método propuesto por Arias (2012); que incluye los siguientes pasos:

1. Búsqueda de fuentes: impresas y electrónicas.
2. Lectura inicial de los documentos disponibles.
3. Elaboración del esquema preliminar.
4. Recolección de datos mediante lectura evaluativa y elaboración de resúmenes.
5. Análisis e interpretación de la información recolectada en función del esquema preliminar.
6. Formulación del esquema definitivo y desarrollo de los capítulos.
7. Redacción de la introducción y conclusiones.
8. Revisión y presentación del informe final.

## **3. Estado del arte**

### **3. 1 Discapacidad motriz y Tecnología**

En el contexto de la presente investigación se busca dar a conocer la importancia del desarrollo tecnológico en el la psicomotricidad de personas con alguna discapacidad en su movimiento corporal, para lo cual se pretende desarrollar algorítmicamente máquinas capaces de brindar calidad de vida inclusiva, eficiente y de fácil acceso. La inactividad motriz de las personas se debe a múltiples factores entre los cuales se destacan las falencias del sistema nervioso, en particular por la fisiología del cerebro o por efecto de traumas (Martínez, Alcaraz, J. R. Cárdenas Valdez, & López, 2018).

*... “Con el manejo masivo de los datos y los avances de la Inteligencia Artificial será posible convertir los sistemas de salud tradicionales en aparatos masivos de prevención y manejo oportuno de las enfermedades y los riesgos de salud. De esta forma, se podrán optimizar sus finanzas y, sobre todo, elevar los indicadores de morbilidad y mortalidad.” ... (El Espectador,2018).*

Avanzar en tecnología convierte a las personas en seres excepcionales, en un mundo con vista al futuro, no se puede dejar atrás el desarrollo mental y que este se base en la inclusión de todos los seres humanos, con la capacidad de realizar acciones bajo cualquier circunstancia.

Los actores del sector desde la academia, integrantes del Consejo Departamental de Ciencia, Tecnología e Innovación hasta jóvenes investigadores del Departamento de Nariño, se encuentran realizando bajo la creación de ideas innovadoras, avances tecnológicos en nuestra jurisdicción (Gobernación de Nariño, 2016).

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

En el Departamento de Nariño ... *“el tema de Ciencia, tecnología e Innovación lo concebimos desde una mirada más amplia, donde no solo se toca el tema del emprendimiento o tecnológico, sino que le estamos apostando a la gente”* ... (Paola Coral, 2016).

## **4. Marco teórico**

En esta etapa se identificaron las distintas temáticas asociadas a la generación y reconocimiento de voz, haciendo un paralelo entre los procesos humanos y tecnológicos. Se inicia con la producción de la voz, explicando la fisiología y el funcionamiento del aparato respiratorio, fonador y resonador. Se resalta la voz como una característica propia de la especie humana, capaz de producir sonidos entendibles y autónomos (*como una huella digital*). A continuación, se estudia el oído humano, explicando su anatomía y el proceso de percepción acústica, el cual se divide en: captación y transformación de la onda mecánica, la conversión de la señal de impulsos eléctricos y la etapa de procesamiento neuronal, en la cual los impulsos son interpretados por el sistema nervioso. Finalmente, se explican las características de las ondas de sonido y la estructura para el reconocimiento de comandos de voz (pre énfasis, segmentación, inventariado y extracción de características), donde se incluyen los métodos HMM (Modelos Ocultos de Markov), reconocimiento usando solo audio y reconocimiento del habla basado solo en fonemas.

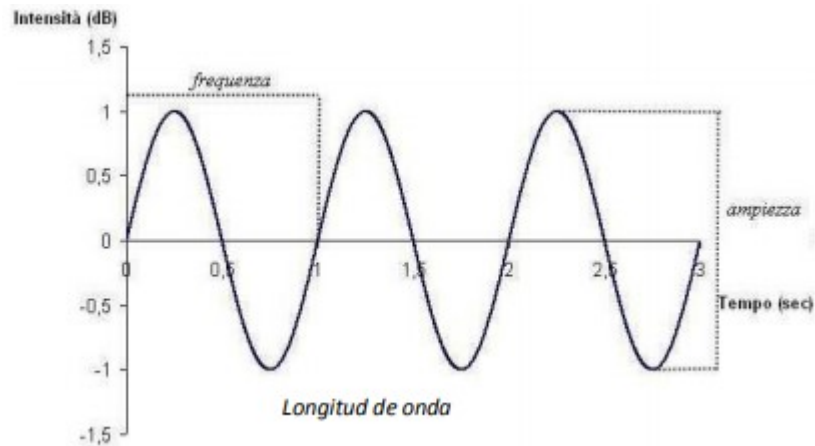
### **4.1 Fisiología de la voz**

La voz humana es el producto de la vibración de las partículas del aire las cuales producen sonido, el cual se emite bajo perturbaciones en forma de onda sonora, con una serie de características entre las cuales se encuentra, amplitud, la longitud de onda, y la frecuencia. La amplitud indica el volumen, medida en decibelios (dB), concepto importante al igual que la frecuencia del sonido, la cual se representa en ciclos por unidad de tiempo y se expresa en hertzios (Hz). (Gonzalez, 2014)

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

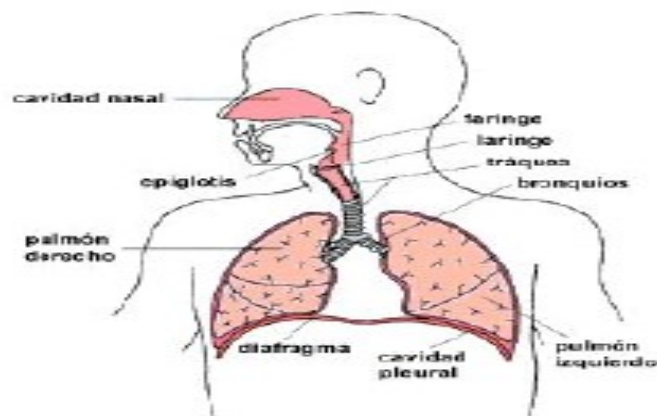
**Figura 1:** Representación de las ondas



Tomado de (Gonzalez, 2014)

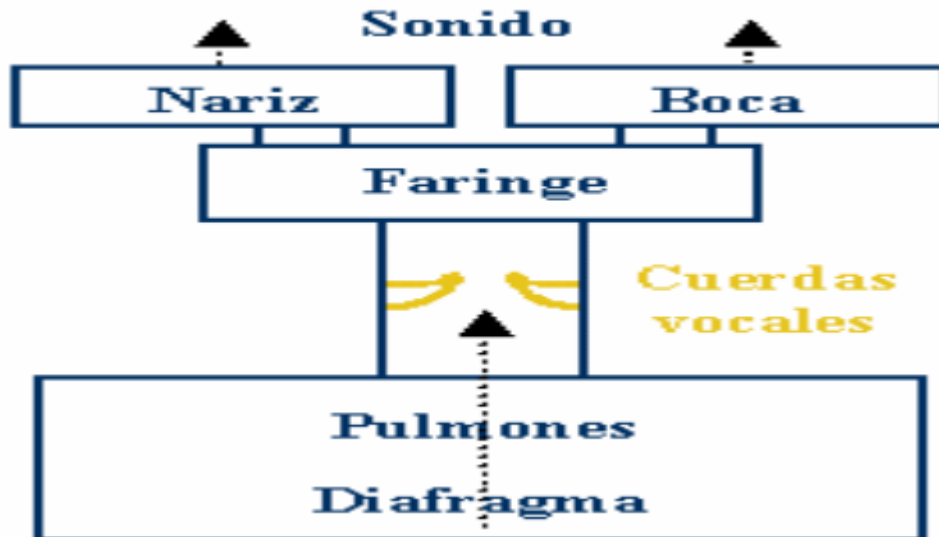
En las Figuras 2,3 y 4, se puede observar el instrumento fonatorio, sus partes en secuencia y las transiciones que surgen.

**Figura 2:** Aparato fonador



Tomado de (Trejos, 2007)

**Figura 3:** Esquema aparato fonador



Tomado de (Voz Humana, 2005)

## 4.2 Clasificación de los Sonidos

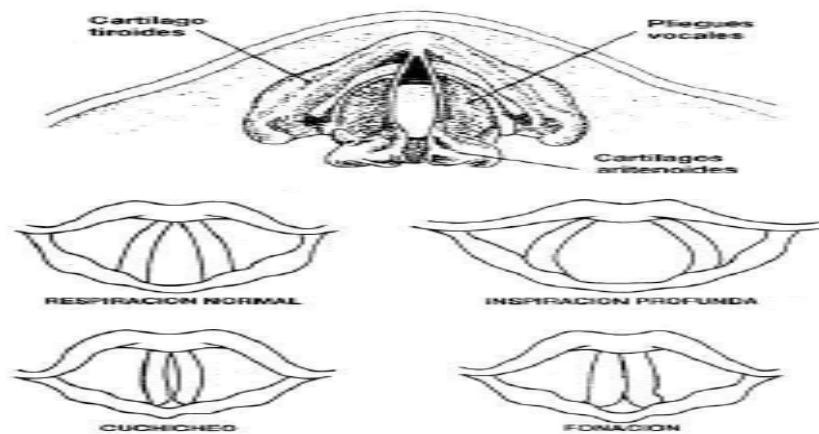
La voz humana es el producto de la acción conjunta de varias estructuras anatómicas, para su estudio se dividen en tres regiones llamadas fuelles, vibrador y resonadores. Los fuelles comprenden todas las estructuras del aparato respiratorio situadas por debajo de la glotis, o infragloticas, especialmente los pulmones y la caja torácica. El vibrador, formado por la laringe como principal órgano de la voz, contiene los pliegues vocales. Los resonadores son todas las cavidades del aparato respiratorio y digestivo que se encuentran por encima de la glotis, también llamadas estructuras supragloticas. Los articuladores son los elementos que participan en la modulación del sonido, para generar los diferentes sonidos del habla humana y son fundamentalmente tres, la lengua, los dientes, y los labios. (Gonzalez, 2014).

### 4.3 Fisiología de la Fonación

Las cuerdas vocales son repliegues o labios en número de cuatro: dos repliegues superiores que son las cuerdas falsas o bandas ventriculares, y dos repliegues inferiores que son las verdaderas cuerdas vocales. Entre estas existe una hendidura o espacio vacío que los limita denominado glotis. Los dos repliegues inferiores, son los que producen el sonido: (Trejos, 2007)

1. Si dichas cuerdas se aproximan y vibran se origina un “sonido sonoro”, pero si no vibran será un “sonido sordo”.
2. La vibración provoca una onda sonora o tono fundamental y unos armónicos que filtrados (en la cavidad bucal y en la nasal) producen un timbre del sonido.
3. Al pasar el aire hacia las cuerdas vocales con mayor o menos energía se produce la intensidad de voz.
4. La duración se produce por un impulso psicomotriz a través del nervio recurrente hacia el diafragma. Este comprime los pulmones el tiempo necesario para la duración deseada.

**Figura 4:** Niveles de apertura glotal

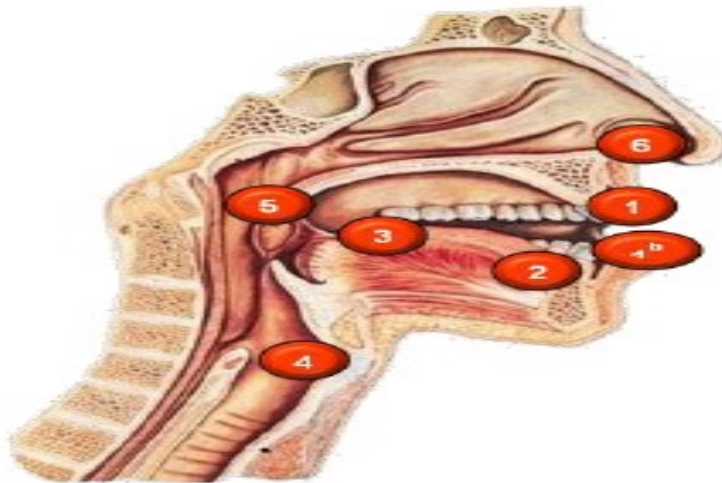


Tomado de ( Gonzalez, 2014)

En la fonación, existe una frase “prefonatoria” antes de la emisión de la voz, en que los pliegues vocales adoptan la denominada “posición fonatoria”, que consiste en aproximarse entre sí, desplazándose cada uno hacia la línea media, con lo que se estrecha el tracto respiratorio a nivel de la glotis. Estando las cuerdas vocales en aducción, el flujo de aire espirado produce la vibración sonora, consistente en la transformación de la energía aerodinámica en energía acústica. ( Gonzalez, 2014)

Como se puede observar en la Figura 7, el punto número 1 está situado en los labios, el número 2 en el ápex de la lengua y la parte interior de los incisivos y las encías, el número 3 en el dorso de la lengua y el paladar, el número 4 en las cuerdas vocales, el número 5 en el velo del paladar y la parte superior de la nasofaringe, y, por último, el número 6 en las narinas. Los sonidos que conforma el habla se pueden producir por tres mecanismos escape, explosión y vibración. ( Gonzalez, 2014)

**Figura 5:** Puntos de articulación



Tomado de ( Gonzalez, 2014)

Las principales unidades fundamentales del sonido, son las vocales y consonantes, las primeras son producidas cuando el tracto vocal se excita por pulsos de aire emitidos por las cuerdas vocales, su vibración es periodica a diferencia de las consonantes donde las

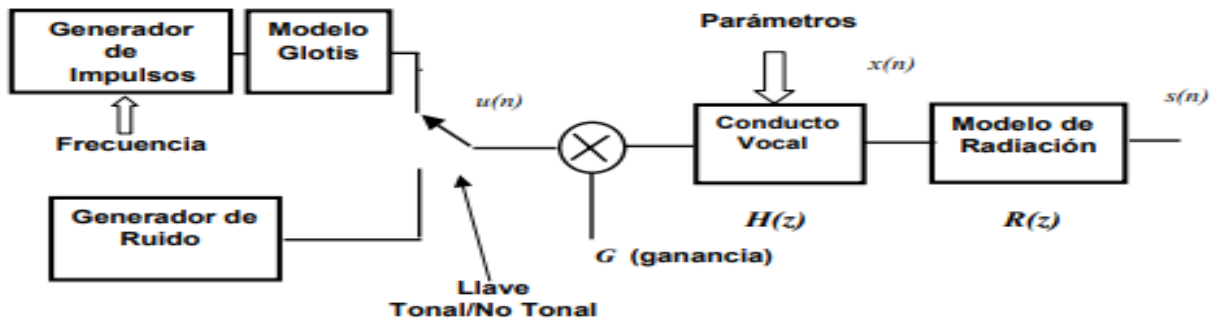


Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

cuerdas vocales están relajadas, Basado en (Gómez, Simancas, Acosta, Meléndez, & Vélez, 2017). Para poder sintetizar la voz de forma artificial se necesita el siguiente modelo de producción de voz.

**Figura 6:** Modelo de producción de voz



Tomado de (Gómez , Rabiner, L. & Juang, B-H, 1998)

## 4.4 Oído humano

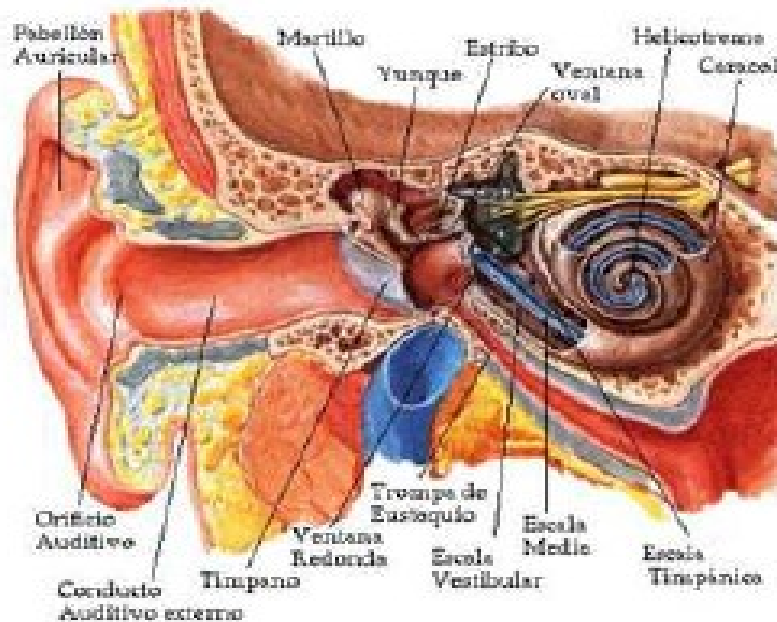
El mecanismo físico de la percepción del habla se ha constituido en un medio muy complejo y avanzado. Este proceso se realiza principalmente en tres etapas: la percepción acústica dividida en captación y transformación de la onda mecánica, la conversión de la señal de impulsos eléctricos y la etapa de procesamiento neuronal, en la cual los impulsos son interpretados por el sistema nervioso. Como seres humanos también se poseen limitaciones respecto a lo que se está en capacidad de escuchar o no, bajo condiciones normales. Este límite se compone de las frecuencias de 20 a 20.000 Hz, sin embargo, la intensidad del sonido y la edad son solo algunos de los factores que pueden disminuir los límites y principalmente los extremos del rango. La captación, procesamiento y traducción de los estímulos sonoros se llevan a cabo en el oído, mientras en la etapa de procesamiento neuronal, en la cual se producen variadas sensaciones auditivas, se ubica en el cerebro. Se distinguen dos regiones o partes del sistema auditivo: la región periférica, en la cual los estímulos sonoros conservan su carácter original de ondas mecánicas hasta

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

el momento de su conversión en señales electroquímicas, y la región central, en la cual se transforman dichas señales en impulsos eléctricos y procesos cognitivos. El oído o región periférica se divide en tres zonas, llamadas oído externo, oído medio y oído interno, de acuerdo a su ubicación en el cráneo, como puede verse en la figura 9. Tomado de (Trejos, 2007).

**Figura 7:** Anatomía del oído humano



Tomado de (Trejos, 2007)

Por otro lado, se aprecia que las señales de sonido que percibimos los seres humanos están llenas de ruidos. El sonido se considera como “vibraciones de partículas y sistemas materiales con masa y elasticidad”. Las vibraciones audibles están consideradas en frecuencias de 16Hz y 20kHz por debajo de 16Hz (infrasonidos) y por encima de 20kHz (ultrasonidos). No es posible para el oído humano captar estas señales, la mayor parte de los sonidos que el ser humano puede captar se encuentran entre los 500 y 8000Hz. Otro de los parámetros que define lo que se puede percibir es la intensidad de sonido, donde

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

podemos situar un rango de 0 a 120Db, los valores cercanos a los 120Db empiezan a generar molestias o dolor en los oídos (Angiono et al., 2017).

El oído está localizado en el peñasco del hueso temporal y contiene estructuras óseas, nervios y músculos, además de aire. Sus principales funciones son la audición y la ayuda en el control del equilibrio. Se divide en tres porciones independientes, cada una de las cuales cumple funciones específicas:

- Oído externo: Compuesto por una porción fibrocartilaginosa lateral y otra ósea medial, capta y transmite el sonido hasta la membrana timpánica.
- Oído medio: Contiene la cadena osicular, su función es amplificar la onda sonora y transmitir las ondas del sonido de un medio aéreo a uno líquido.
- Oído interno: Compuesto por los aparatos coclear y vestibular, sus principales funciones son transformar las ondas acústicas en impulsos nerviosos y a través del vestíbulo ayudar a mantener el equilibrio. (Revista médica Sanitas, 2015)

## 4.5 Sistemas para la adquisición de señales

Diferentes autores han identificado la necesidad de representar parámetros físicos como, por ejemplo: la temperatura, humedad, presión, caudal, etc., de forma cuantificada; a partir de la aplicación de diferentes métodos descriptivos. Muchos avances tecnológicos necesitan de entradas de señal analógica, las cuales deben ser transformarlas en señales digitales para su posterior procesamiento y análisis. El resultado de esto será una toma de decisiones, que puede conllevar a acciones físicas. Las señales analógicas se encuentran en el entorno físico y se han representado de diferentes formas y bajo diferentes métodos, para obtener una descripción lo más parecida a ellas. La rama de la instrumentación trata de percibir esa realidad utilizando diferentes métodos para la extracción de información del entorno mediante la identificación y cuantificación de las características físicas suficientes para describir diferentes estados que se presentan cotidianamente (Ochoa, 2018). En nuestro caso las señales analógicas a identificar son señales de acústicas y en específico aquellas generadas por la voz humana. (Inguilan, 2019)

### 4.5.1 Características de las ondas de sonido

Según (Lopez, 2018) la onda de sonido es un tipo de onda mecánica de desplazamiento, la que se propaga de forma longitudinal transportando energía, a través de materiales elásticos o viscosos ya sea sólido líquido o gaseoso, el ondas de sonido se propagan a través del aire y hacen vibrar sus partículas ocasionando cambios de presión y densidad en dirección de las mismas, la medida adoptada para cuantificar estas ondas es la presión sonora, sus unidades son newtons por metro cuadrado ( $N/m^2$ ), existe una relación con la presión atmosférica representada como:

$$p(t) = P(t) - P(atm) \quad \text{Ecuación 1}$$

Donde  $t$  es el tiempo,  $p(t)$  la presión sonora en el instante de tiempo  $t$ ,  $P(t)$  la presión momentánea del aire en el instante de tiempo  $t$  y  $P(atm)$  la presión atmosférica.

Otra característica de las ondas es la potencia la cual se entiende como la cantidad de energía por unidad de tiempo, la potencia se mide en un punto fijo (intensidad sonora o densidad de potencia), sobre un área infinitesimal de tamaño  $dA$

$$I = \frac{dP}{dA} \quad \text{Ecuación 2}$$

Donde  $dP$  es la potencia acústica en el área  $dA$  detectada por el sensor estará perpendicular a la dirección de la onda sonora, la intensidad sonora se mide en watts por metro cuadrado ( $W/m^2$ ).

## 4.6 Estructura para el reconocimiento de comandos de VOZ

### 4.6.1 Pre-procesamiento

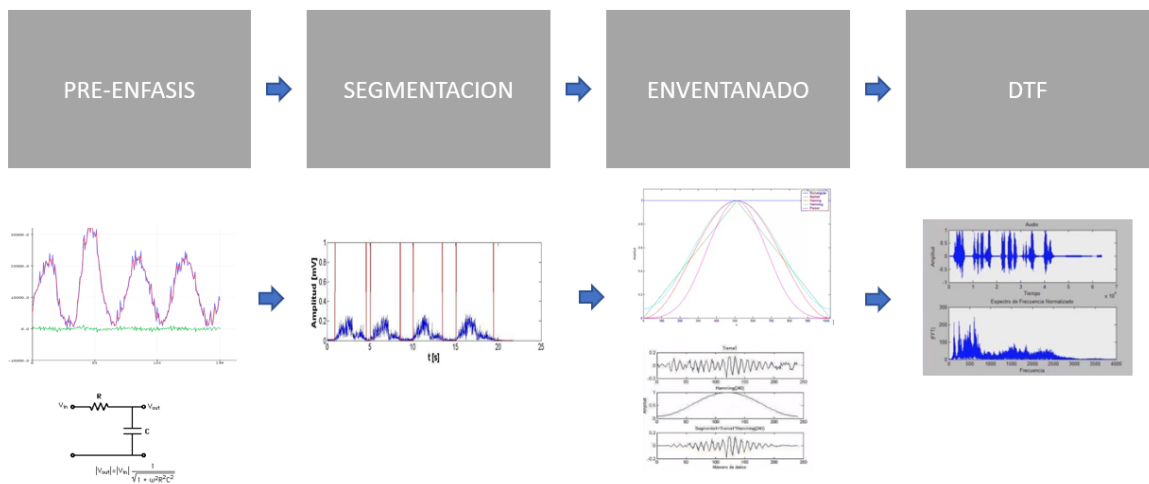
La señal de salida de un sistema en general se debe procesar de una forma adecuada para la siguiente etapa de la operación. La señal puede ser por ejemplo,

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

amplificada; podría contener interferencias que eliminar; ser análoga y requerir su digitalización o viceversa; requerir un cambio de voltaje etc. Los elementos que conectan un sistema electrónico con su entorno muchas veces no están preparados para ser conectados con el núcleo del sistema, las etapas de acondicionamiento de señal, hacen compatibles dichas conexiones. En este sentido las muestras tomadas necesitan de un pre-procesamiento de sus señales, con el fin de extraer los datos más importantes, la estructura interna del pre procesamiento se muestra en la Figura 11.

**Figura 8:** Estructura interna de pre procesamiento



Basado en (matlab,2019)

#### 4.6.2 Pre énfasis

Los acondicionadores de señal consisten en circuitos que transforman los parámetros eléctricos de salida de los transductores en una señal eléctrica (corriente, voltaje o frecuencia), la cual puede medirse de forma factible. Este proceso incluye las transformaciones de las señales eléctricas resultantes del transductor previo al

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

procesamiento de extracción de la información deseada instrumentación electrónica: transductores y acondicionadores de señal (Miguel & Bolad, 2015, pag 6-7).

Razones del condicionamiento de señales

- Cuando la señal eléctrica medida no está dada en magnitud o intensidad, por lo cual es conveniente un circuito que realiza esta conversión Convertir una señal en un tipo de señal adecuado. Protección para evitar el daño al siguiente elemento.
- Obtención del nivel adecuado de la señal, en este caso el acondicionamiento implica: aumentar la señal a niveles superiores al ruido electrónico, y filtrar señales para eliminar el ruido por interferencia eléctrica.
- Manipulación de la señal. Por ejemplo, convertir una variable en una función lineal, mediante circuitos específicos o programas de cálculo adecuados.

Se utiliza un filtro de alto orden para enfatizar previamente la muestra de voz (Kavitha, Nachammai, Ranjani, & Shifali, 2014).

$$Y(d) = X(d) - C * X(d-1) \quad 0-1$$

Donde  $X(d)$  es la señal de entrada,  $Y(d)$  es la señal de salida y  $C$  es una constante con un valor entre 0.9-1

### 4.6.3 Segmentación

En la segmentación se divide el vector de la señal acústica en tramos de entre 20 a 40 ms, para realizar el análisis de cada tramo de forma individual. La frecuencia para este tipo de señales es comúnmente 8 kHz o mayor. Se puede distribuir estos cuadros de análisis de modo que se realice una superposición a fin de no perder información entre los bordes de inicio y final de cada cuadro. Uno de estos métodos se le denomina ventana de Hamming; donde la superposición vuelve a colocar las características de la señal en las derivadas al tamaño del cuadro, generalmente en las potencias de dos.

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

#### 4.6.4 Enventanado

Proceso en el que se integran los cuadros para evitar distorsiones y discontinuidades en las partes inicial y final de cada cuadro. Cuando se usa la transformada de Fourier en estas muestras de señales acústicas se debe realizar las superposiciones a fin de que las transiciones de señal sean lo más suaves posible y facilite el inicio y termino de cada uno, comúnmente se utiliza para el reconocimiento de comandos la ventana de Hamming  $H(d)$ .

$$H(d) = 0.54 - 0.46 \cos\left(\frac{2\pi d}{D-1}\right)$$

$$0 \leq d \leq D-1$$

La señal procesada se obtiene por la relación de:

$$Y(d) = X(d) * H(d)$$

Aquí  $D$  = número de muestras en cada cuadro,  $Y(d)$  = señal de salida,  $X(d)$  = señal de entrada

#### 4.6.5 Extracción de características

Para la extracción de propiedades o características de la señal, se destacan los siguientes parámetros:

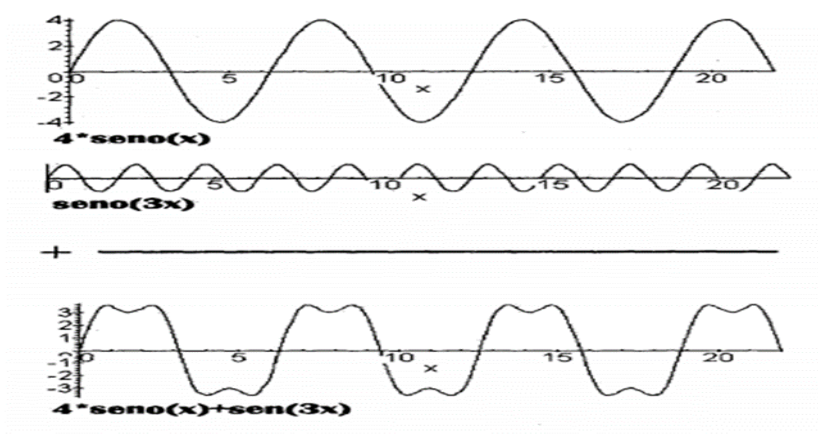
- Propiedades espectrales – energía espectral
- Tono y Timbre
- Análisis por bancos de filtros digitales
- Componentes vectoriales formantes
- Valores Temporales-inicio y finalización de la señal de audio
- Valores temporales-inicio y finalización de la señal de audio; amplitudes, entre otros.
- Transformada discreta de Fourier (DTF)
- Análisis Cepstral

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

Cada uno de los anteriores parámetros deben ser extraídos por cada segmento pronunciado. Ahora bien, entre las técnicas más destacables para análisis espectral se tiene desde algoritmos clásicos como la FFT (Transformada rápida de Fourier), hasta modelos regresivos y autorregresivos, Modelos Predicción Línea, entre otros. (Camargo,2010)

**Figura 9.** Descomposición de una señal compleja en sumatorio de señales simples



Tomado de (J Bobadilla et al.,1999).

- **Transformada de Fourier**

Las series trigonométricas surgieron en la Matemática en el siglo XVIII, en relación con el estudio de las pequeñas oscilaciones de medios elásticos. Más precisamente, con el siguiente problema que consiguió la atención y el esfuerzo de numerosos físicos y matemáticos del momento.

En la actualidad, la teoría de series de Fourier puede presentarse usando los conceptos y métodos de Análisis Funcional. Más concretamente está íntimamente relacionado con la Integral de Lebesgue, los espacios de Hilbert (extensión a dimensión infinita de espacio



Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

euclídeo) y el espacio de funciones cuadrado integrable. Fue Hilbert quien identificó una función dada  $f$  con sus coeficientes de Fourier  $\{f_n, n \in \mathbb{N}\}$ .

Estos coeficientes satisfacen, en general, la condición  $\sum_{n=1}^{\infty} |f_n|^2 < +\infty$  (si  $f$  es de cuadrado integrable). Hilbert introdujo además el espacio  $l^2$  de sucesiones de números reales  $a_n$  tales que la serie  $\sum_{n=1}^{\infty} |a_n|^2$  es convergente. Posteriormente, Riesz y Fischer demostraron la existencia de una aplicación biyectiva entre el conjunto de las funciones de cuadrado integrable  $L^2(a, b)$  (en un intervalo finito dado  $(a, b)$ ) y el conjunto  $l^2$  (a cada función se le hace corresponder sus coeficientes de Fourier). (Trinidad, 2017)

Se define la Transformada de Fourier de  $f$  como la función  $F: \mathbb{R} \rightarrow \mathbb{C}$ , dada por:

$$\mathcal{F}\{f(t)\} = F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\omega t} dt.$$

Típicamente,  $f(t)$  se conoce como función de la variable del tiempo y  $F(\omega)$  como función de la variable frecuencia. Además, un símbolo minúsculo representa una función de tiempo y las transformadas de Fourier de este tiempo están representadas por el mismo símbolo en mayúscula como función de la frecuencia. La definición varía según los gustos en la aparición de ciertas constantes: el exponente puede ser  $-i\omega t$ ; la fórmula de inversión (2.1) lleva un factor  $1/2\pi$  multiplicativo a la integral, para recuperar la simetría.

En general, la Transformada de Fourier es una función compleja:

$$F(\omega) = R(\omega) + iI(\omega) = |F(\omega)|e^{i\theta(\omega)}$$

Donde:

$R(\omega)$  es la parte real de la FT

$I(\omega)$  es la parte imaginaria de la FT

$|F(\omega)|$  es la amplitud del espectro de Fourier de  $f(t)$  que está dada por  $\sqrt{R^2(\omega) + I^2(\omega)}$  y  $\theta(\omega)$  es el ángulo de fase de la transformada de Fourier y está dada por  $\tan^{-1} [I(\omega) / R(\omega)]$ . (Trinidad, 2017)

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

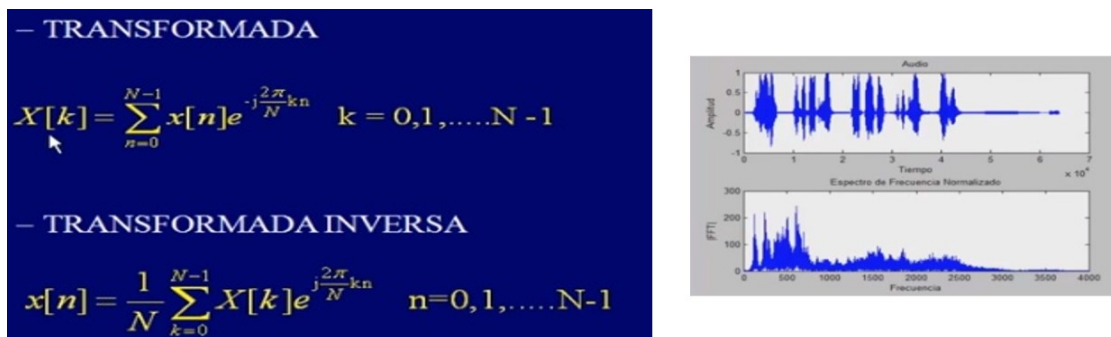
La transformada de Inversa de Fourier (**TIF**)

Si  $f$  y  $F$  son integrables en sentido Lebesgue, entonces:

$$\mathcal{F}^{-1}\{F(\omega)\} = f(t) = \int_{-\infty}^{\infty} F(\omega)e^{2\pi i\omega t} d\omega.$$

El término de la derecha es una función continua en  $\omega$ , de modo que  $f$  coincide en casi todo punto con una función continua y se da la igualdad en los puntos de continuidad de  $f$ . A  $f$  se le llama Transformada Inversa (o anti transformada) de Fourier de  $F$ . (Trinidad, 2017).

**Figura 10.** Transformada de Fourier



Tomado de (Matlab, 2019)

La transformada de Fourier se utiliza para descomponer la señal en sus respectivas componentes en el dominio de la frecuencia y los predictores lineales buscan establecer sistemas que emulen y describan el comportamiento del trato vocal. (Trejos et al,2007)

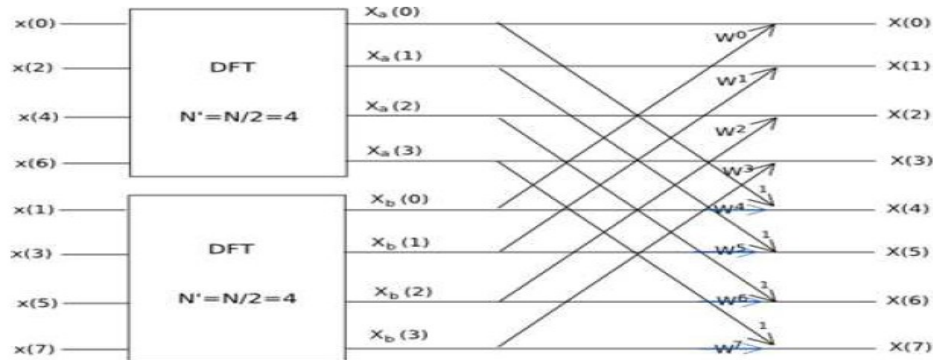
La transformada rápida de Fourier (FFT) resulta de gran importancia por su capacidad para operar en tiempo discreto y la obtención de algoritmos que permitieran reducir el número de operaciones necesarias para obtener el área electrónica (reduce el número de tracciones en la obtención de la Transformada de Fourier conocida este método como "Decimación en el Tiempo")

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

$$X(k) = X_a(k) + W^k X_b(k)$$

A partir de esta expresión se puede obtener una representación grafica de la transformada rápida de Fourier para 8 términos. (Trejos et al,2007)

**Figura 11.** Diagrama FFT para N=8



Tomado de (Trejos et al,2007)

- **Banco de filtros de Mel**

Se toman grupos que contengan un periodo y se suman para obtener una idea de energía, presente en cada región de las frecuencias (Ceballos, Serna-morales, Prieto, Gómez, & Redarce, 2011). Para reconocimiento voz solo son necesarios los primeros 12-13 coeficientes denominados coeficientes cepstrales en las frecuencias de Mel (Lyons).

Imitando la forma en que los humanos escuchan, la escala de frecuencia Mel, tiene un esparcimiento lineal de frecuencia por debajo de los 1000 Hz, las señales de voz tienen más energía en las frecuencias más bajas, las siguientes formulas son para calcular los mels de una frecuencia dada en Hz.

$$mel(f) = 2595 \cdot \log\left(1 + \frac{f}{700}\right)$$

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

Para cada tono con una frecuencia actual  $f$  Hz, un tono subjetivo se mide en la escala de Mel. El pitch de un tono de 1 kHz, 40 dB por encima de la audiencia perceptual se conoce entonces como 1000 mels. En este caso sería:

$$mel(f) = 2595 \cdot \log\left(1 + \frac{f}{700}\right) = 181.312.111,042623$$

### • Transformada de cosenos discreta de Fourier

Se obtienen los coeficientes evaluados que representan la señal de voz normalizada.  $N$ , es el número de filtros triangulares y  $m_j$  son los coeficientes a la salida del banco de filtros.

$$C_i = \sqrt{\frac{2}{N}} * \sum_{j=1}^N m_j * \cos\left(\frac{\pi}{N} * (j - 0.5)\right)$$

El tamaño de la matriz transformada discreta del Coseno es  $N*N$ , consecuentemente los índices  $k$  y  $m$  tomaran valores desde 0 hasta  $N-1$ , de la manera análoga a como ocurre con la transformada de Fourier.

## 4.6.6. Reconocimiento de voz

Al proceso de convertir una señal acústica a una secuencia de palabras en forma de texto por medio de un dispositivo de control se le llama reconocimiento de voz. Según el artículo de Sistema audiovisual para reconocimiento de comandos (Ceballos et al., 2011) existen diversas formas de aplicar el reconocimiento de voz, algunos de los más usados son los HMM (Modelos Ocultos de Markov), reconocimiento usando solo audio y reconocimiento del habla basado solo en fonemas.

**Figura 12:** Estructura interna de procesamiento



Basado en (Gordillo, 2018)

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

- **DTW Dinámico Time Warping**

Técnica DTW es utilizada para determinar y comparar las distancias entre una y otra curva de señales, el resultado de esta presenta un modelado no bien definido, razón por la que en estos momentos no es tan empleada, requiere de segmentos básicos de palabras, usado para reconocimientos aislados, el procesamiento aumenta en cuanto aumenta la base de datos de modelos de entrenamiento, convirtiéndose en una desventaja en el momento de aumentar la unidad de reconocimiento (Ureña, 2011).

- **Modelos ocultos de Markov (HMMs)**

Es un método de reconocimiento estadístico de señales el cual es empleado comúnmente. Los HMMs son modelos con secuencias estadísticas en los que se procesan señales estacionarias, estos representan una secuencia de tiempo y variación de la magnitud de voz, un HMM describe una palabra como una transición estados en un solo sentido, para más de 20 estados hay un crecimiento logarítmico de la probabilidad (Ceballos et al., 2011).

Estos fonemas recolectados estarán acompañados de su correspondiente representación como serie de palabras, cuanto más información de modelo se tenga el reconocimiento será más exacto, los modelos de lenguaje tendrán información sobre las palabras y las posibles combinaciones que se deben realizar, para esto se requiere de gran cantidad de datos, por lo que se opta por usar aproximaciones como las basadas en N-Gramas asignando probabilidad a posibles próximas palabras, así estos se podrían usar para probabilidades de frases enteras, estos términos provienen de los modelos de Markov. A esto se relacionan las características de señal del hablante, como pueden ser estilo, tono y ritmo del habla, las características son únicas para cada persona ya que el tracto vocal y fisiología son diferentes al comparar sus variaciones de frecuencia, se debe filtrar las señales de ruido ya que son principal causa que impide la identificación de las palabras (Gil Vásquez, Castillo Ossa, & Flórez Hurtado, 2017).

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

- **Reconocimiento de voz registro de sonidos**

El proceso se fundamenta en la vectorización del audio, donde dicha señal se convierte en una sucesión de valores, donde cada uno representa un vector de propiedades, donde se describe de forma detallada el sonido o fonema registrado. (Camargo,2010)

- **Reconocimiento del habla basado en fonemas**

Teniendo en cuenta que se debe crear una base de datos a partir del procesamiento, reconocimiento, análisis acústico, frecuencia, amplitud, estructura armónica y resonancia de la voz entre otros.

La forma acústica de un fonema depende fuertemente del contexto acústico en el que sucede. A este efecto se le llama coarticulación. Debido a su presencia se debe realizar una percepción auditiva, la cual consiste en la codificación de cambios de todos, la muestra y la cuantización para obtener una muestra de señal codificada dependiendo del entorno.

La forma de hablar, el vocabulario, la gramática y el entorno deben presentar estudios profundos y su metodología debe conllevar a la producción y solución que demuestre un procesamiento computalizado que permita llevar a cabo el dominio y funcionamiento de los comandos de voz. (Trejos et al,2007)

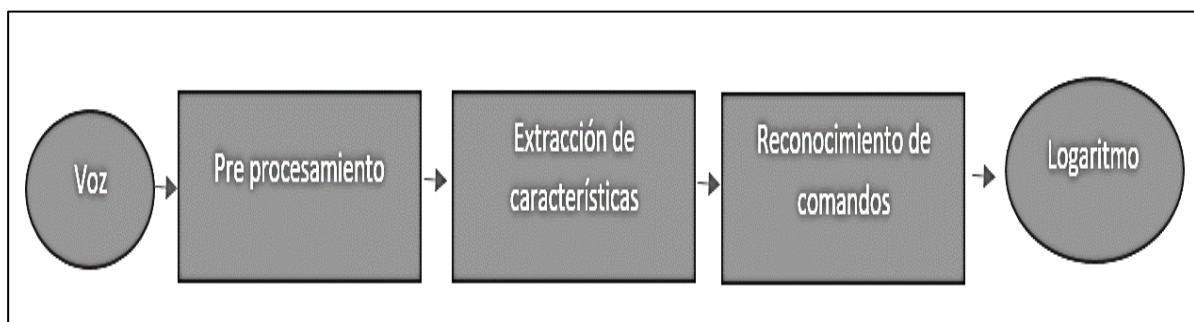
## 5. Resultados

### 5.1 Estructura de un sistema de reconocimiento de voz para silla de ruedas

Esta investigación ha conllevado a definir una estructura para el reconocimiento de comandos de voz más común y con los elementos necesarios sin dejar de ser efectivos para ser aplicados en una silla de ruedas.

Lo anterior se basa en la implementación de un algoritmo con algunas tareas específicas como la realización del prototipo (*el cual en este caso no desarrollará*), se da en el entrenamiento de codebook u oculto de markov de cada palabra y reconocimiento de palabras aisladas (Iván Higuera Martínez, Espitia Flórez, Méndez Medina, 2015).

**Figura 13:** Estructura de un sistema de reconocimiento de voz



Basado en (Gordillo, 2018)

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

## 5.1.1 Pre procesamiento

- **Preénfasis**

Se utiliza un filtro de alto orden para enfatizar previamente la muestra de voz.

$$Y(d) = X(d) - C * X(d-1)$$

Donde  $X(d)$  es la señal de entrada,  $Y(d)$  es la señal de salida y  $C$  es una constante con un valor entre 0.9-1.

Luego se filtran las señales inferiores a 20Hz y superiores a 20kHz.

- **Segmentación**

En la segmentación para nuestro caso se aplica lo comúnmente usado y que nos conviene utilizar serán cuadros de 25ms y de 8kHz.

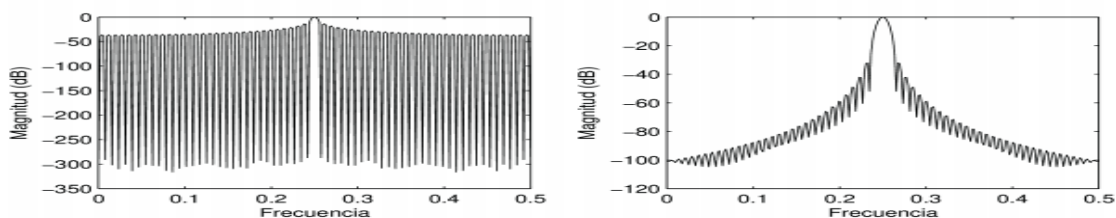
Las muestras segmentadas en su nivel fonético, se realiza por medio de conjuntos de unidades de palabra, donde los modelos ocultos de Markov (HMMs), modelan cada una de estas unidades, donde se corrige el error que pueda presentarse.

- **Enventanado**

Para este caso se aplica las ventanas de Hamming las cuales llevaran a cabo un realce de las características de los segmentos supe posicionados para evitar la pérdida de datos en los bordes de cada 25ms como se ha escogido y 8000Hz.

Producto de esta señal, se obtiene una convolución del espectro de la señal con espectro de ventana.

**Figura 14:** Estructura de un sistema de reconocimiento de voz



Basado Matlab



Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

### 5.1.2 Extracción de características

Se ha escogido para la extracción de características los modelos HMMs, ya que nos ofrece buenos resultados de comparación y es de menos complejidad.

“Para lograr dicho resultado, la señal de voz ingresa a un módulo de procesamiento de señales en el que se extraen los vectores de características sobresalientes que son enviados posteriormente al decodificador” (L.J. Gil , L.F. Castillo, R.D. Flórez, 2016).

Los HMMs y las ANNs presentan habilidades complementarias para el reconocimiento automático del habla, lo que se refleja en las prestaciones superiores de los sistemas híbridos (Solera Ureña, 2011).

Figura 15: Extracción de características

Extracción de características de la voz humana transformar audio en parámetros como timbre, duración e intensidad

$Mel(f) = 2595 \cdot \log_{10} \left( 1 + \frac{f}{700} \right)$

**ECUACIÓN DEL CALCULO DE ENERGÍA**

$$S_t[m] = \ln \left( \sum_{n=0}^{N-1} |K_t[m]|^2 H_n[m] \right), 0 \leq m \leq M$$

**Transformada Discreta Coseno (DCT)**

$$F(u) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \cos \left( \frac{(2u+1)m\pi}{2N} \right) f(x)$$

**Definición de HMM (cont.)**

Un HMM es un modelo  $M = (Q, \Sigma, \pi, A, B)$ , donde:

- $Q$  es un conjunto finito de **estados** (que incluye uno **final**,  $F$ )
- $\Sigma$  es un conjunto finito de **símbolos** o **alfabeto**
- $\pi \in [0, 1]^Q$  es un vector de **probabilidades iniciales**
- $A \in [0, 1]^{Q \times Q}$  es una matriz de **probabilidades de transición**
- $B \in [0, 1]^{Q \times \Sigma}$  es una matriz de **probabilidades de emisión**

**Ejemplo:**

$a : 0.5 \quad a : 0.2$   
 $b : 0.5 \quad b : 0.8$

$Q = \{0, 1, F\}$   
 $\Sigma = \{a, b\}$

$\pi$	0	1
	0.6	0.4

$A$	0	1	$F$
0	0.3	0.7	
1		0.6	0.4

$B$	$a$	$b$
0	0.5	0.5
1	0.2	0.8

Basado en (Inguilan,2019)

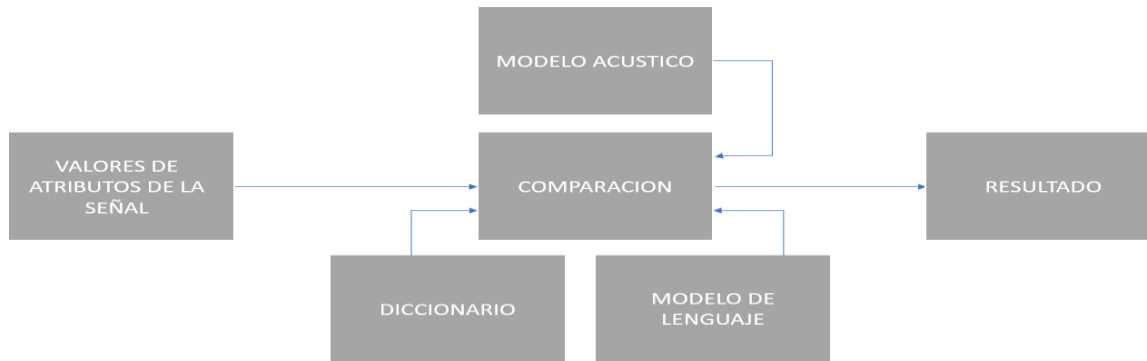
### 5.1.3 Reconocimiento de voz

Para solucionar el reconocimiento de voz en palabras aisladas hemos considerado la implementación de modelos ocultos de Markov debido a la baja complejidad ideal para aplicar a esta unidad de identificación (palabras), además de esta técnica valoramos el uso de descripciones estadísticas que mejora la capacidad del sistema y reduce los errores en comparación con otros métodos como lo es DTW.

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

**Figura 16:** Estructura Final de un modelo de reconocimiento de voz



Basado en (Inguilan,2019)

## 5.2 Elaboración de un algoritmo para reconocimiento de comandos de voz

*“Los algoritmos genéticos son métodos adaptativos de búsqueda y optimización. Están basados en la evolución genética de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza acorde con los principios de la selección natural y la supervivencia de los más fuertes. Tomando como base esta idea, los algoritmos genéticos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones lleva a valores óptimos del problema, que depende en buena medida de una adecuada configuración del algoritmo”* (Guevara Maldonado, 2018)

### 5.2.1 Redacción

Entrada: Señal de voz

Salidas: Comandos de voz

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

1. Captura de señal de voz.  $X(t)$
2. Digitalización  $X(n)$ 
  - 2.1. Muestreo: es discreta la señal con un periodo  $t$
  - 2.2. Cuantificación: dar un valor  $n$  a la amplitud de la señal muestreada
  - 2.3. Codificación: utilizar un número de bits para representar cada uno de los valores cuantificados
3. Filtros de pre-énfasis (filtro pasa bajas)  $(1-\alpha Z^{-1})$  donde  $0.7 < \alpha < 0.9$
4. Segmentación: dividir en cuadros de 25ms a 8kHz con 10ms de superposición.
5. Enventanado: aplicar ventana de Hamming para realce de la parte central del fonema  
 $x[n]$ : señal de voz  
 $w[n]$ : ventana de análisis  
 $N$ : Tamaño de ventana  
 $M$ : Desplazamiento  
 $W[M-n]$   $w[2M-n]$   $w[3M-n]$
6. Extracción de características MFCC
  - 6.1. Aplicar Transformada de Fourier a cada segmento
  - 6.2. Pasar cada segmento por el banco de filtros de Mel
  - 6.3. Aplicar logaritmos de la energía:
    - La energía de la señal varía en el tiempo:  
 Fonemas sordos menos energía que fonemas sonoros  
 Consonantes sonoras menor energía que vocales
    - La Energía de tiempo corto pone de manifiesto estas variaciones:  

$$E_n \sum_{m=-\infty}^{\infty} (x(m)w(n-m))^2 \quad E_n \sum_{m=-\infty}^{\infty} x^2(m) \cdot h(n-m)$$
    - Magnitud promedio de tiempo corto:  
 Es una medida alternativa a la energía de tiempo corto  
 Es menos sensible a la amplitud de las muestras  

$$M_n \sum_{m=-\infty}^{\infty} |x(m)|w(n-m)$$
7. Crear modelo acústico basado en Módulos ocultos de Markov HMM
8. Crear diccionario de los principales comandos de voz
  1. Adelante = [ a ðe 'laN te ]
  2. Atrás = [ 'a tras ]
  3. Derecha = [ de 're tja ]
  4. Izquierda = [ iθ 'kjer ða ]
  5. Pare = [ 'pa re ]
9. Enteramiento de comandos de voz.
10. Test de reconocimiento.

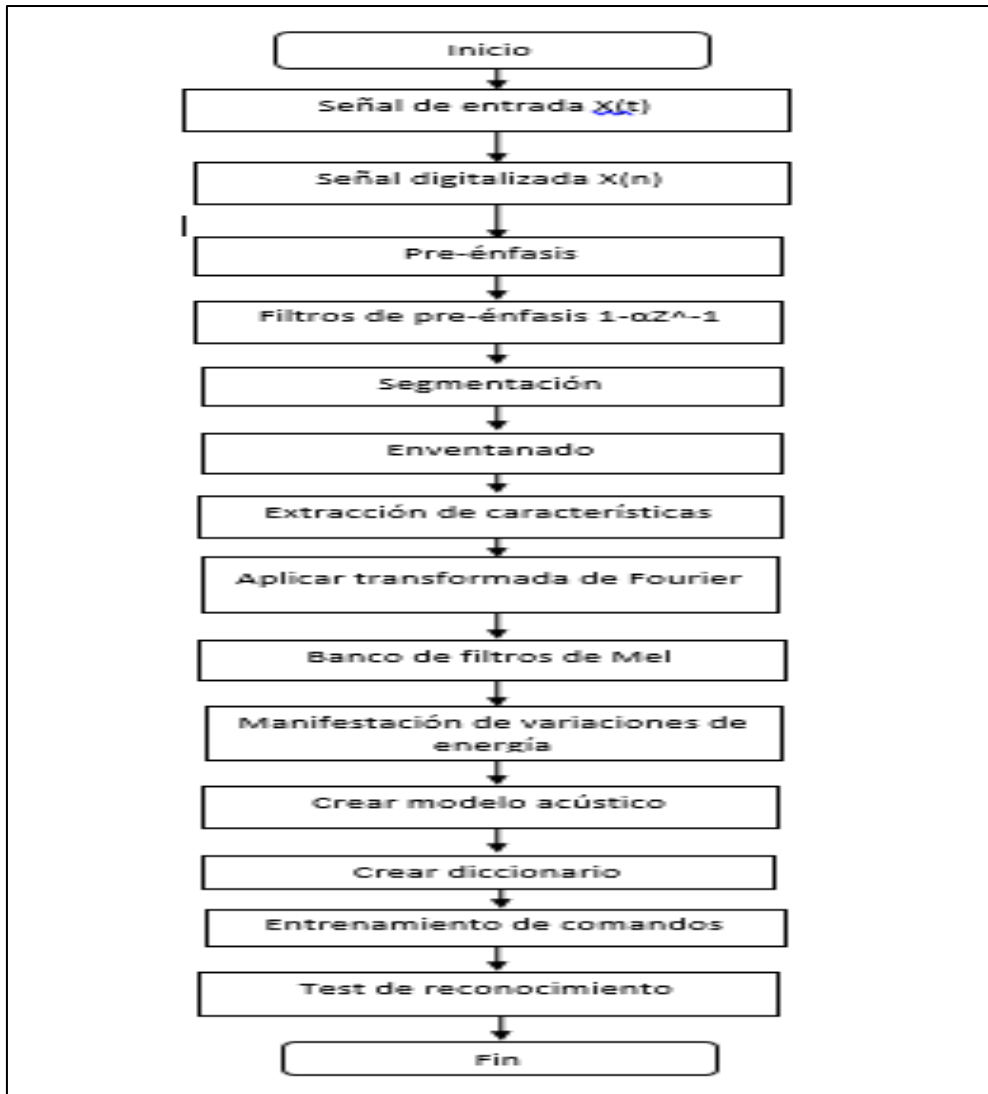
Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

### 5.3 Diagrama de flujo

Figura 17: Diagrama de flujo

Se desarrolló uno de los diagramas más importantes como lo es el diagrama de flujo de datos el mismo que indica los procesos que tiene el sistema. (Narváez Huera Luis, 2014)



## **6. Conclusiones y recomendaciones**

### **6.1 Conclusiones**

La investigación basada en comandos de voz, permitió interiorizar en el diseño de algoritmos con el fin de que se adapte a cualquier medio electrónico como es el caso de una silla de ruedas, por medio del análisis de investigaciones y proyectos de grado, he logrado reconocer el sistema de procesamiento de señales y las implicaciones físicas, lógicas y matemáticas que conllevan a un perfecto funcionamiento de un algoritmo.

Por medio de la investigación de procesamiento de señales se coadyuva a las personas discapacitadas en la solución de problemas que comprometen su movilidad y motricidad, logrando con esto que se integren a la sociedad y desarrollen funciones acordes a sus capacidades.

De igual forma pude reconocer que la voz del ser humano esta conformada de diferentes y únicas características las cuales puedes ser demostradas bajo condiciones matemáticas y electrónicas que nos conllevan a su procesamiento, entrenamiento y resultado, es así como se logra implementar esta serie de procesos en la ayuda para la misma humanidad y su libre locomoción y desarrollo inclusivo.

Teniendo en cuenta que el aprendizaje a futuro será retroalimentarse y crear nuevas formas de procesar la voz, que coadyuven mediante métodos, equipos, sensores y guías con el fin de que la medicina siga un cauce de investigación electrónico que ayude a las personas con discapacidad a mejorar su calidad de vida, y que la paraplejia no sea un

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

obstáculo, ya que por medio de las creaciones tecnológicas grandes personalidades a nivel mundial han demostrado que la tecnología es un aliado de un futuro sin obstrucción alguna.

Aplicar los modelos ocultos de Markov, en el análisis de señales de voz, constituyo en mi investigación una forma esencial y prioritaria, ya que es una herramienta donde el algoritmo conlleva a la posibilidad de obtener segmentos de voz ante un complejo y determinado procesamiento.

## 6.2 Recomendaciones

Que se desarrollen proyectos a nivel mundial, en los cuales se podría innovar en el futuro para emprender investigaciones similares o fortalecer la investigación realiza y coadyuven al mejoramiento e impulso de las personas con algún tipo de discapacidad motriz.

Continuar con la investigación en procesamiento de señales de voz, con el fin de que se innoven en nuevas prácticas tecnologías que permitan a futuro desaparecer todo estigma y obstáculo para personas discapacitadas.

que se implemente el presente algoritmo, y se llevado a pruebas en Matlab y posteriormente en una estructura plena e identificable como lo es la silla de ruedas, donde se lleve a cabo base de datos con comandos, tipo, adelante, atrás, izquierda, derecha, en condiciones normales y entrenamiento de los comandos, con el fin de conceptuar resultados el preprocesamiento, y analizar todas sus señales.

Crear bases de dado, e implementarlas en trabajos futuros, y que estas sean dadas en todas las condiciones, llevar a la estructura total, sensores de obstáculos con el fin de que las personas puedan sentirse seguras de que los obstáculos pueden ser reconocidos, y no generen impedimento alguno en su movilización.

## Bibliografía

Arias, F. (2012).

El Proyecto de Investigación. Introducción a la metodología científica. Venezuela: Editorial EPISTEME.

Colegio Santa Elena (2015)

Guía de estudio El Sonido

Angiono, V., Thompson, M. A. M., Lucini, M. B., Serra, M., & Serra, S. (2017).

Fonoaudiología, bases de la comunicación humana. *Journal of Visual Languages & Computing* (Vol. 1).

Sarah Lewin (2015)

El Origen del Oído.

Hernando Pericas (1993)

Técnicas de Procesado y representación de la señal de voz para el reconocimiento del habla en ambientes ruidosos.

Paredes Cruz (1997)

Mejoramiento de imágenes Monocromáticas usando la transformada de Fourier discreta del coseno.

J Bobadilla, Gómez, Bernal (1999)

La Transformada de Fourier. Una Visión Pedagógica

Ceballos, A., Serna-morales, A. F., Prieto, F., Gómez, J. B., & Redarce, T. (2011).

Sistema audiovisual para reconocimiento de comandos Audiovisual system for recognition of commands, *19*, 278–291.

Ernesto Cutillas (2016)

Distribución mundial de la población con discapacidades en relación con los

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

patrones geograficos del desarrollo humano.

Gil Vásquez, L. J., Castillo Ossa, L. F., & Flórez Hurtado, R. D. (2017). Reconocimiento de comandos de voz en español orientado al control de una silla de ruedas. *Revista UIS Ingenierías*, 15(2), 35–48. <https://doi.org/10.18273/revuin.v15n2-2016003>

Gómez, J., Simancas, J., Acosta, M., Meléndez, F., & Vélez, J. (2016). Algoritmo de reconocimiento de comandos voz basado en técnicas no-lineales. *Espacios*, 38(17), 18. Retrieved from <http://repositorio.cuc.edu.co/xmlui/handle/11323/904>

Alvarez Gloria (2008)

Modelos Ocultos de Markov

Camargo J. (2010)

Sistema de reconocimiento de voz humana por hardware

Ministerio de Salud(2017)

Ley Estatutaria 1618 del 2013 " Por medio de la cual se establecen las disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad"

Sorge Radovani L (2015)

Los problemas vocales del actor en formación

Gordillo, C. D. A. (2018). *Realce e Reconhecimento de Voz Contínua em Ambientes Adversos*.

Dosal Gonzalez R (2014)

Producción de la Voz y el Habla, La Fonación

Fernandez Baillo R (2011)

Revisión de los modelos de producción de voz después de una laringectomía total: opciones de una laringectomía total: opciones de calidad de voz.

Trejos Posada H (2007)

Motor computacional de reconocimiento de voz: Principios básicos para su construcción

Preparadores de Oposiciones para la enseñanza (2005)

La voz humana y su fisiología. Clasificación de las voces. La voz en la adolescencia: Características y problemática.

Hernandez LAdino N, Lara Lozada M (2013)

Propuesta pedagógica de sensibilización para el manejo adecuado de la voz que



Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

minimiza los riesgos de enfermedad y mejora el desempeño laboral de los docentes.

Munar E, Rosselló J, Mas C, Morente P, Quetgles M, (2002)

El desarrollo de la audición humana .

Guamán, M. S. C. (2018). Medición del umbral de audición en bajas frecuencias e infrasonido, 95. Retrieved from

<http://dspace.udla.edu.ec/bitstream/33000/10205/1/UDLA-EC-TISA-2018-21.pdf>

F.A Sepúlveda (2004)

Extracción de Parámetros de Señales de Voz usando técnicas de análisis tiempo-frecuencia.

Trinidad Flor (2017)

Transformada de Fourier y su aplicación en procesamiento digital de imágenes.

Informe de Secretaría -OMS. (2014). Proyecto de acción mundial de la OMS sobre discapacidad 2014-2021: Mejor salud para todas las personas con discapacidad, 27. <https://doi.org/EB134/2014/REC/2>

Kavitha, R., Nachammai, N., Ranjani, R., & Shifali, J. (2014). Speech Based Voice Recognition System for Natural Language Processing, 5(4), 5301–5305. Retrieved from

<https://pdfs.semanticscholar.org/14f6/6f7aaebb56f9b7eb7a008d84afce6708bc12.pdf>

Lopez, M. A. (2018). Análisis experimental de la factibilidad del uso de compresión con pérdida en la clasificación automática de sonidos biológicos, (June). Retrieved from [https://www.researchgate.net/profile/Manuel\\_Aguilera4/publication/327447358\\_Analisis\\_experimental\\_de\\_la\\_factibilidad\\_del\\_uso\\_de\\_compresion\\_con\\_perdida\\_en\\_la\\_clasificacion\\_automatica\\_de\\_sonidos\\_biologicos/links/5b900b5945851540d1cc2369/Analisis-experimenta](https://www.researchgate.net/profile/Manuel_Aguilera4/publication/327447358_Analisis_experimental_de_la_factibilidad_del_uso_de_compresion_con_perdida_en_la_clasificacion_automatica_de_sonidos_biologicos/links/5b900b5945851540d1cc2369/Analisis-experimenta)

Coral P. (2016)

<https://xn--nario-rt.a.gov.co/inicio/index.php/sala-de-prensa/noticias/164-narino-por-la-humanizacion-de-la-tecnologia>.

Maldonado, C. B. G. (2018). *Desarrollo de algoritmos eficientes para identificación de usuarios en accesos informáticos*. UNIVERSIDAD COMPLUTENSE DE MADRID FACULTAD. Retrieved from <https://eprints.ucm.es/46037/1/T39510.pdf>

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

- Martí, J. (2017). La investigación - Acción participativa estructura y fases, 37–41.  
[https://doi.org/10.1577/1548-8659\(2001\)130](https://doi.org/10.1577/1548-8659(2001)130)
- Martínez, A. M., Alcaraz, G. E. V., J. R. Cárdenas Valdez, & López, C. E. V. (2018). Manejo de Silla de Ruedas Eléctrica por Comandos de Voz Personalizado, 6. Retrieved from  
<http://fcqi.tij.uabc.mx/usuarios/revistaaristas/numeros/N12/articulos/150-155.pdf>
- Miguel, M. G., & Bolad, E. M. (2015). *Instrumentacion Electronica: Transductores y acondicionadores de señal*. Retrieved from  
<http://www.editorial.unican.es/libro/instrumentacion-electronica-transductores-y-acondicionadores-de-senal>
- Ochoa, B. joosely L. (2018). Analisis de consumo monofasico. Retrieved from  
[https://addi.ehu.es/bitstream/handle/10810/29158/TFG\\_BerthaLorenzoOchoa.pdf?sequence=1&isAllowed=y](https://addi.ehu.es/bitstream/handle/10810/29158/TFG_BerthaLorenzoOchoa.pdf?sequence=1&isAllowed=y)
- Pérez, E., Pocerros, F., & Jose, V. (2013). Sistema de Seguridad Por Reconocimiento de Voz, 74. Retrieved from  
[http://tesis.bnct.ipn.mx/dspace/bitstream/123456789/12309/1/Sistema de Seguridad por Reconocimiento de Voz \(Tesis de Ingenieria ESIME\).pdf](http://tesis.bnct.ipn.mx/dspace/bitstream/123456789/12309/1/Sistema%20de%20Seguridad%20por%20Reconocimiento%20de%20Voz%20(Tesis%20de%20Ingenieria%20ESIME).pdf)
- Ureña, R. S. (2011). *Maquinas de vectores soporte para reconocimiento robusto de habla*. Retrieved from [https://e-archivo.uc3m.es/bitstream/handle/10016/12577/Tesis\\_Ruben\\_Solera\\_Urena.pdf](https://e-archivo.uc3m.es/bitstream/handle/10016/12577/Tesis_Ruben_Solera_Urena.pdf)
- Marco Stalin Carrillo Guaman (2018)  
Medición del Umbral de audición en bajas frecuencias e infrasonido.
- Gómez Durán Julieth, Simancas García José, Acosta Coll Melisa, Meléndez pertuz Farid, Vélez-Zapata Jaime (2018).  
Algoritmo de reconocimiento de comandos de voz, basados en técnicas no lineales.
- Luis Alberto Valente Vela (2017)  
Reconocedor de palabras aisladas mediante la voz.
- Narváez Huera Luis (2014)  
Implementación de un prototipo de sillas de ruedas automatizada con comandos de sonido para mejoramiento de la enseñanza en el Manejo y orientación a personas con limitaciones físicas.

Estudio del procesamiento de comandos de voz empleando los modelos ocultos de Markov, aplicado al control autónomo de una silla de ruedas.

---

Rebeca Hernández (2016)

Sistema de control activado por voz para uso en Domótica

Higuera Martínez, Espitia Flórez, Méndez Medina (2015)

prototipo de silla de ruedas comandada por voz, empleando HMM en un ambiente controlado.

L.J. Gil , L.F. Castillo, R.D. Flórez, (2016)

reconocimiento de comandos de voz en español orientado al control de una silla de ruedas.

Solera Ureña (2011)

Máquinas de vectores soporte para reconocimiento robusto del habla.

Cesar Byron Guevara Maldonado (2018)

Desarrollo de algoritmos eficientes para identificación de usuarios en accesos informáticos.

Pérez Eyra, Poceros Fernando, Villalobos José (2013)

Sistema de seguridad por reconocimiento de voz.

Kasse Fawaz, Huan Feng, Kang Geun Shin (2017)

System and method for authenticating voice commands for a voice assistant.

Kavitha, Nachammai , Ranjani, Shifali (2014)

Speech Based Voice Recognition System For Natural Language Processing.

Rabiner, L. & Juang, B-H. (1998)

Speech Recognition by Machine, Chap. 47 in

The Digital Signal Processing Handbook, CRC Press, IEEE Press.

Rabiner, L. & Juang, B-H (1993).

*Fundamentals of Speech Recognition*, Prentice Hall, N.J.