



# **UNIVERSIDAD JOSÉ ANTONIO PÁEZ**

## **SISTEMA DE SEGURIDAD POR RECONOCIMIENTO DE VOZ**

### **Autores:**

Izquierdo, María

C.I.: 25.981.156

Maldonado, Alejandro

C.I.: 25.049.335

Urb. Yuma II, calle N° 3. Municipio San Diego  
Teléfono: (0241) 8714240 (master)



**REPÚBLICA BOLIVARIANA DE VENEZUELA  
UNIVERSIDAD JOSÉ ANTONIO PÁEZ  
FACULTAD DE INGENIERÍA  
ESCUELA DE TELECOMUNICACIONES**

**SISTEMA DE SEGURIDAD POR RECONOCIMIENTO DE VOZ**

**Trabajo de grado presentado como requisito para optar al título de  
INGENIERO EN TELECOMUNICACIONES.**

**Autores:**

Izquierdo, María

C.I.: 25.981.156

Maldonado, Alejandro

C.I.: 25.049.335

**Tutor:**

Ing. Blanco, Rainier



**FI-T-005-2019-ICR**

Valencia, 21 de Marzo de 2019

Ciudadano:  
Maria Izquierdo  
C.I: 25.981.156  
Alejandro Maldonado  
C.I: 25.049.335  
Presente-

Cumplo con informarle que la Comisión de Trabajo de Grado y Pasantías de la Facultad de Ingeniería en su reunión N° 01-2019 de fecha 21-03-2019 aprobó el proyecto de trabajo de grado titulado **SISTEMA DE SEGURIDAD POR RECONOCIMIENTO DE VOZ**. Presentado por usted(es) como requisitos para optar al título de Ingeniero en Telecomunicaciones

Se ratifica la designación del Ing. Rainier Blanco, C.I: 11.556.607 y la Ing. Alicia De Pizzella, C.I: 4.598.880 como Tutores Académicos que lo asesorarán en el desarrollo de este proyecto.

Atentamente,



**Prof. Luis Lira**  
Decano de la Facultad de Ingeniería

c.c. Coordinación de Pasantías y Trabajo de Grado (1).  
LH/c.



REPÚBLICA BOLIVARIANA DE VENEZUELA  
UNIVERSIDAD JOSÉ ANTONIO PÁEZ  
FACULTAD DE INGENIERÍA  
ESCUELA DE TELECOMUNICACIONES

#### ACEPTACION DEL TUTOR

Quien suscribe, Ingeniero Rainier Blanco, titular de la cedula de identidad N° 11.556.607, en mi carácter de tutor del trabajo de grado presentado por los ciudadanos IZQUIERDO F. MARÍA V. titular de la cedula de identidad N° 25.981.156 y MALDONADO C. ALEJANDRO portador de la cedula de identidad N° 25.049.335 titulado "**SISTEMA DE SEGURIDAD POR RECONOCIMIENTO DE VOZ**". presentado como requisito parcial para optar al título de ingeniero en Telecomunicaciones, considero que dicho trabajo reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del jurado examinador que se designe.

En San Diego, a los 13 DIAS DEL MES DE MAYO del año 2019

Ing. Rainier Blanco

C.I: 11.556.607

## ÍNDICE GENERAL

CONTENIDO	Pp.
<b>INDICE DE FIGURAS.....</b>	<b>viii</b>
<b>INDICE DE TABLAS.....</b>	<b>ix</b>
<b>RESUMEN INFORMATIVO.....</b>	<b>x</b>
<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPÍTULO</b>	
<b>I EL PROBLEMA</b>	
1.1 Planteamiento del problema. ....	3
1.2 Formulación del problema.....	4
1.3 Objetivos de la investigación.....	5
1.3.1 Objetivo general.....	5
1.3.2 Objetivos específicos.....	5
1.4 Justificación del problema.....	5
1.5 Alcance.....	6
<b>II MARCO TEÓRICO</b>	
2.1 Antecedentes.....	7
2.2 Bases teóricas.....	8
2.2.1 Sistema Biométrico.....	8
2.2.2 Reconocimiento de voz.....	9
2.2.3 Historia del reconocedor de voz.....	10
2.2.4 Etapas de un sistema de identidad biométrico.....	12

2.2.5 Características de la señal de voz: .....	14
2.2.6 Señal acústica.....	17
2.2.7 Percepción auditiva.....	20
2.2.8 Micrófono. ....	20
2.2.9 Muestreo y Cuantización. ....	22
2.2.10 MATLAB. ....	25
2.2.11 Obtención de información mediante el micrófono. ....	26
2.2.11 Métodos de reconocimiento de voz. ....	26
2.3 Definición de términos básicos. ....	28

### **III MARCO METODOLÓGICO**

3.1 Tipo de investigación. ....	31
3.2 Diseño de la investigación.....	31
3.3 Nivel de la investigación. ....	32
3.4 Técnicas e instrumentos de recolección de datos .....	32
3.4.1 Técnicas de recolección de datos.....	32
3.4.2 Instrumentos de recolección de datos. ....	33
3.5 Fases de la investigación. ....	33

### **IV RESULTADOS**

4.1 Fase I: Caracterizar la voz humana mediante métodos acústicos, matemáticos y físicos.....	35
4.2 Fase II: Clasificar los datos obtenidos para poder crear una base de datos y utilizarlos en el sistema de seguridad. ....	40
4.2.1 Creación de la base de datos .....	41
4.3 Fase III: Desarrollar un programa con lenguaje de alto nivel que funcione como interfaz de usuario para el reconocimiento de la voz. ....	44
4.3.1 Diseño de los componentes (botones y ejes) que formarán la GUI.....	44

4.3.2 Programación de cada uno de los componentes ante la interacción del usuario.....	46
4.4 Fase IV: Evaluar el programa de reconocimiento de voz mediante la simulación. ....	52
4.4.1 Pruebas del reconocimiento.....	54
4.4.2 Pruebas con personas no registradas.....	57

## **V CONCLUSIONES Y RECOMENDACIONES**

5.1 Conclusiones .....	60
5.2 Recomendaciones .....	61
<b>REFERENCIAS .....</b>	<b>62</b>

## INDICE DE FIGURAS

Figura 1. Proceso de comunicación del habla.....	15
Figura 2. Órganos del cuerpo que intervienen en el aparato fonador. ....	16
Figura 3. Señal senoidal muestreada.....	23
Figura 4. Diagrama de generación de la señal acústica .....	35
Figura 5. Diagrama de la adquisición de voz.....	36
Figura 6. Señal tratada digitalmente para ser muestreada.....	37
Figura 7. Diagrama de la digitalización de la señal de voz.....	38
Figura 8. Diagrama de los filtros de énfasis.....	39
Figura 9. Diagrama de segmentación, ventaneo y recorte de la señal acústica .....	40
Figura 10. Diagrama que explica los procesos de extracción de características.....	40
Figura 11. Interfaz principal de MATLAB. ....	41
Figura 12. Programación de la base de datos.....	42
Figura 13. Programación del promedio de cada usuario.....	43
Figura 14. Base de datos de usuarios. ....	43
Figura 15. Modo de acceso a GUIDE en MATLAB. ....	44
Figura 16. GUIDE Quick Start. ....	45
Figura 17. Espacio de trabajo para crear una interfaz.....	45
Figura 18. Diseño de la interfaz gráfica.....	46
Figura 19. Programación para la captación de voz. ....	47
Figura 20. Programación de la señal limpia y grafica a mostrar.....	47
Figura 21. Interfaz gráfica luego de haber presionado el botón de grabar.....	48
Figura 22. Programación de los LPC. ....	48
Figura 23. Programación de distancias. ....	49
Figura 24. Programación de contadores.....	49
Figura 25. Programación de la identificación. ....	50
Figura 26. Interfaz gráfica luego de presionar identificar.....	50
Figura 27. Interfaz gráfica cuando no se es identificado. ....	51
Figura 28. Programación del botón de reiniciar.....	51
Figura 29. Programación del botón de salir. ....	51
Figura 30. Interfaz gráfica luego de presionar salir. ....	52
Figura 31. Programación de la gráfica del programa.....	52

## INDICE DE TABLAS

Tabla 1. Pruebas con nivel de ruido de 47 a 55 dBA.....	53
Tabla 2. Pruebas con nivel de ruido de 63 a 70 dBA.....	53
Tabla 3. Pruebas con nivel de ruido de 75,6 a 80 dBA.....	54
Tabla 4. Resultados intentando acceder como María.....	54
Tabla 5. Resultados intentando acceder como Alejandro.....	55
Tabla 6. Resultados intentando acceder como Virginia.....	56
Tabla 7. Resultados para un hombre adulto.....	57
Tabla 8. Resultados para una chica adolescente.....	58



**REPÚBLICA BOLIVARIANA DE VENEZUELA  
UNIVERSIDAD JOSÉ ANTONIO PÁEZ  
FACULTAD DE INGENIERÍA  
ESCUELA DE TELECOMUNICACIONES**

**SISTEMA DE SEGURIDAD POR RECONOCIMIENTO DE VOZ**

**Autores:** María Izquierdo y Alejandro Maldonado

**Tutor:** Ing. Rainier Blanco

**Fecha:** Junio del 2019.

**RESUMEN INFORMATIVO**

El presente proyecto de investigación surge a raíz de la problemática, que enfrentan todas las personas por la inseguridad vivida actualmente en el país, por otra parte, es de hacer notar que no todos los sistemas de seguridad implementados en casi todos los hogares, son eficientes y confiables a la hora de presentarse un intruso. En consecuencia, el proyecto de investigación tiene como principal objetivo, el diseño de un sistema de seguridad por reconocimiento de voz, el cual pueda entregarle al usuario la seguridad, confiabilidad y eficiencia necesaria para lograr resguardar tanto sus bienes como la integridad física de las personas que se encuentran dentro del recinto, de esta manera también se busca que este trabajo de investigación sirva como base para futuras investigaciones y mejoramiento en los sistemas de seguridad biométricos.

Descriptores: Sistemas de seguridad, reconocimiento de voz, MATLAB.

## INTRODUCCIÓN

Una de las amenazas más peligrosas para la seguridad es la suplantación, en la cual alguien se hace pasar por otra persona. Los sistemas de seguridad enfrentan esta amenaza mediante el reconocimiento y la autenticación. El reconocimiento es el mecanismo mediante el cual se identifica a un individuo específico entre muchos, mientras que la autenticación se ocupa de verificar la identidad de ese individuo. El sistema de reconocimiento y autenticación del usuario se puede clasificar en base a lo que él sabe (una contraseña), en base a lo que el posee (una tarjeta inteligente) o en base a sus características humanas (biométricas).

En la presente investigación se tiene como objetivo analizar cómo se puede desarrollar un sistema de seguridad confiable y eficiente, de acuerdo a lo antes mencionado se realizaron estudios acerca de los sistemas de seguridad por reconocimiento de voz con un tipo de investigación documental para la línea de diseño y desarrollo. En atención a lo planteado se puede decir que, el sistema le va a permitir al usuario poder grabar una palabra por medio de un micrófono y luego la información pasará por un proceso de verificación para así comenzar la comparación con una base de datos ya existente. En este sentido la señal de voz es ingresada a la computadora, grabada y procesada por los algoritmos del programa, obteniendo los parámetros LPC significativos de la señal de voz y posteriormente ser almacenados en la computadora. Es importante resaltar que la selección de reconocimiento permite que la palabra sea comparada con la base de datos almacenada en la computadora en un proceso llamado comparación de coeficientes LPC.

El presente trabajo de investigación está estructurado en cuatro capítulos, con el fin de cumplir las normativas establecidas por la Universidad José Antonio Páez, dichos capítulos se describen a continuación:

Capítulo I: este capítulo va referido al problema, su planteamiento el cual se trata de comprobar durante el desarrollo del proyecto por medio de los objetivos generales y específicos, así como la justificación del estudio y su alcance.

Capitulo II: se centra en los antecedentes, trabajos realizados los cuales aportan datos necesarios para llevar a cabo nuestro proyecto, así como en las bases teóricas, las cuales son necesarias para el desarrollo de la investigación y en último caso tenemos la definición de los términos básicos.

Capitulo III: este capítulo, el marco metodológico describe el tipo y diseño de la investigación a utilizar para cumplir los objetivos del proyecto, además se detallan las fases metodológicas que se utilizaran para desarrollar el proyecto.

Capitulo IV: en este capítulo, se habla de los recursos utilizados para poder realizar este proyecto.

## **CAPÍTULO I EL PROBLEMA**

### **1.1 Planteamiento del problema.**

En los últimos años en Venezuela debido a la inestabilidad política que se vive se han deteriorado los principios básicos de los pobladores de este país, es decir que por una mala política la nación se enfrenta a una crisis económica dejando así a miles de venezolanos en la pobreza extrema. El Tiempo, Mercado (2018).

De acuerdo con esto se genera una gran descomposición social lo cual fomenta las acciones que van en contra de la seguridad de otros. Desde que se recuerda, la inseguridad vivida en Venezuela es uno de los principales problemas que aqueja a la población. Las cifras semanales de personas muertas en las principales ciudades, así como las interminables noticias de secuestros, robos, hurtos, violaciones y decenas de delitos más, son parte de la vida cotidiana. En Venezuela, inexplicablemente y contrario a lo que es una sostenida tendencia a nivel mundial, el Estado desde hace varios años, no facilita información frecuente y detallada sobre los indicadores del delito.

En cuanto a la violencia, según el ranking 2016, de las 50 ciudades más violentas del mundo, elaborado anualmente por el Consejo Ciudadano para la Seguridad Pública y Justicia Penal A.C. (México), en Venezuela se encuentran siete de esas urbes. Caracas es el número 1 a nivel mundial, y está acompañada de Maturín, Ciudad Guayana, Valencia, Barquisimeto, Cumaná y Barcelona. Estas ciudades manejan indicadores que van desde los 47 hasta los 130 fallecidos cada 100 mil habitantes. En este sentido en el estado Carabobo los tres municipios más violentos están alrededor del lago de Valencia y son fronterizos con el estado Aragua los cuales son Diego Ibarra (Mariara); Carlos Arvelo (Guigue) y los Guayos. Observatorio Venezolano de Violencia, Yuncoza (2017).

Por otra parte, es de hacer notar que los sistemas de seguridad los cuales se implementan habitualmente como circuitos cerrados de cámaras y presencia de personal humano para resguardar instalaciones no son suficientes o simplemente no cumplen con los estándares seguridad por lo cual pueden fallar dejando así a los habitantes que se encuentran dentro del recinto bajo un gran peligro.

Referente a lo mencionado, es necesario tener mejores y más innovadores sistemas de seguridad que interactúen con el propietario de la instalación y que sean capaces de tener un procesamiento de información veloz y precisa que les permita tomar decisiones rápidas y confiables.

Entorno a esto se propone el diseño de un sistema de reconocimiento de voz, por medio de algoritmos implementados en un software para analizar la señal capturada por un micrófono y obtener un correcto reconocimiento de la voz.

Vinculado al concepto, el sistema almacena los datos del usuario básicamente por el micrófono el cual capta la voz del usuario y la convierte en una señal eléctrica luego es detectada, muestreada y cuantizada para ser reconstruida posteriormente como una señal digital la cual va estar representada por una combinación de números binarios es decir “1” y “0”.

Por último, a como se busca mejorar la seguridad del recinto con este sistema, básicamente se guarda la información del usuario, en este caso es la voz, con una serie de parámetros únicos que solo el podrá tener, luego el sistema se implementa en las puertas del recinto para que así solo esa persona tenga acceso al interior. Por otra parte, si algún intruso o persona desconocida intenta entrar al recinto no se le permitirá el acceso ya que no es una persona autorizada en el sistema.

## **1.2 Formulación del problema.**

¿Cómo evitar el ingreso de personas externas a los recintos, a través de un sistema de seguridad de reconocimiento de voz confiable y fácil de utilizar por el usuario?

### **1.3 Objetivos de la investigación**

#### **1.3.1 Objetivo general.**

Diseñar un sistema de seguridad por reconocimiento de voz.

#### **1.3.2 Objetivos específicos.**

- Caracterizar la voz humana mediante métodos acústicos, matemáticos y físicos.
- Clasificar los datos obtenidos para poder crear una base de datos y utilizarlos en el sistema de seguridad.
- Desarrollar un programa con lenguaje de alto nivel que funcione como interfaz de usuario para el reconocimiento de la voz.
- Evaluar el programa de reconocimiento de voz mediante la simulación.

### **1.4 Justificación del problema.**

En relación a la inseguridad actual que se presenta en la ciudad de Valencia se realizará un sistema de seguridad para el hogar utilizando la voz, con la finalidad de proponer a los habitantes de la ciudad un sistema capaz de resguardar su hogar.

Dentro de ese marco, implementar un sistema de seguridad en el hogar por reconocimiento de voz la seguridad será mejor, eficaz y económica a largo plazo. Se hará uso del software MATLAB para desarrollar e implementar el reconocedor de voz, debido a que se está familiarizado con el software, es mejor y más preciso que otros.

Por otra parte, este sistema de seguridad está enfocado principalmente para resguardar a las personas que se encuentran en los recintos urbanísticos para así reducir el número de robos dentro de la misma ya que al tener un sistema de reconocimiento de voz totalmente independiente no es necesario la presencia de vigilantes en el área de entrada, lo cual permite que los mismo realicen rondas de vigilancia constantemente dentro de la urbanización y no se focalice todo el personal de vigilancia a las entradas y salidas de dicha urbanización.

En cuanto al punto de vista metodológico la investigación servirá de base y apoyo para investigaciones futuras, ofreciendo referencia bibliográfica que sirva de consulta o guía ya que representará un conocimiento confiable

### **1.5 Alcance.**

Con la investigación se pretende llegar a crear un diseño capaz de resguardar a los usuarios, ya que, la violencia es un tema bastante común en Venezuela. Asimismo, se busca llegar no solo a la población que se encuentra en el Estado Carabobo sino también a nivel nacional ya que este proyecto está basado en un diseño el cual puede implementarse en cualquier hogar.

## **CAPÍTULO II MARCO TEÓRICO**

### **2.1 Antecedentes.**

A continuación, se realizará una síntesis de los trabajos que han servido como referencia y apoyo para el presente trabajo de investigación, brindándole cuerpo teórico y metodológico. Dichos trabajos son los siguientes:

Camargo, J (2013), egresado de la Universidad Pontificia Bolivariana, llevo a cabo un trabajo de grado bajo el título **“Sistema de reconocimiento de voz por hardware”**, la investigación realiza la descripción de un sistema de Reconocimiento Autónomo de Habla (RAH), el cual permite registrar, validar, identificar y generar acciones a partir de palabras o frases básicas. Específicamente en el caso del proyecto desarrollado, se eligió un sistema basado en hardware (circuitos integrados especializados), los cuales permiten desarrollar aplicaciones stand-alone (independientes).

La información suministrada por la investigación, apporto ideas y técnicas referentes al estudio de la voz, diseño de la programación y verificación de palabras a utilizar en el trabajo de investigación. Por otra parte, permitió el desarrollo e implementación de una base de datos que facilita el desarrollo del proyecto.

Asimismo, González, J (2013), realizó en la Universidad de Granada, como requisito para obtener el doctorado de tecnologías multimedia en el departamento de teoría de la señal, telemática y comunicaciones, un trabajo investigativo titulado **“Reconocimiento de voz con datos perdidos o inciertos”**, dicha investigación adopta un marco de trabajo en el que las características de voz usadas por el reconocedor son procesadas para mitigar la degradación producida por el ruido. Bajo dicho marco genérico, a lo largo del trabajo proponen varias técnicas que, usando estimación bayesiana MMSE e información a priori sobre la voz y/o el ruido,

permiten obtener características más limpia. La información proporcionada por la investigación, usando grabaciones y las técnicas propuestas estiman un conjunto de parámetros que se emplean para realizar las características extraídas de la voz ruidosa, es decir, aporta técnicas para eliminar el ruido en el reconocimiento de voz.

Por último, Caloto, C (2010), realizó en la Universidad Complutense de Madrid como requisito para obtener el título de ingeniero informático, un trabajo de investigación titulado “**Tecnología de reconocimiento por voz y su aplicación en video juegos**”, la investigación consistió en el estudio de los diferentes modelos acústicos, generación de dichos modelos en distintos idiomas, e integración de los mismos en la infraestructura de un juego con el objetivo de manejar el avatar del juego a través de órdenes de voz, sin necesidad de utilizar ninguna interfaz física, como puede ser el teclado o ratón. Ahora bien, al utilizar los Modelos Ocultos de Markov se pueden obtener representaciones numéricas de cada sonido o fonema, este se encarga de recopilar información de distintas fuentes y obtener una representación que resulte aceptable para estas fuentes. De esta manera, si conseguimos varias representaciones de un fonema, podemos analizarlos con los modelos de Markov y obtener una representación que satisfaga el mayor número de individuos.

La información proporcionada por la investigación, permite la descomposición de la voz en diferentes parámetros para su posterior estudio, por otra parte se pueden obtener representaciones numéricas de la voz para realizar cálculos específicos. También es importante destacar que las técnicas utilizadas en esta investigación permiten un mejor desarrollo del algoritmo empleado para el sistema de seguridad.

## **2.2 Bases teóricas.**

### **2.2.1 Sistema Biométrico.**

Un sistema biométrico es un sistema de identificación de personas que se sirve de la biometría informática para condicionar el acceso a un bien o un servicio. Los mecanismos de control automáticos de acceso a bienes o servicios incluyen, además, bases de datos y sistemas físicos como puertas de acceso controladas electrónicamente.

Los aparatos de lectura de huellas dactilares o de análisis de voz son ejemplos comunes de sistemas biométricos.

Asimismo, construye un modelo con la información capturada y un modelo es una aproximación a la realidad. La voz de un individuo es única, pero su registro biométrico podría coincidir con el de otra persona debido a errores en la representación numérica de la información, por ejemplo. Además, cuando los sistemas de seguridad están conectados a redes de cómputo se hace posible alterar la información por medio de programas dañinos, lo que vulnera la seguridad. Normalmente los sistemas biométricos son utilizados en empresas, entes gubernamentales o fuerzas de la ley, para el control de registros del personal, o como medida de seguridad para el acceso a diferentes áreas restringidas, ya que estos rasgos ofrecen una confiabilidad, por ser únicos de cada persona.

### **2.2.2 Reconocimiento de voz.**

El reconocimiento de voz es una tecnología biométrica que utiliza la voz de un individuo para lograr su identificación. Este tipo de soluciones biométricas son bastantes populares, dada la cantidad de dispositivos que pueden usarse para tomar las muestras de voz y su facilidad de integración. Hay que tener en cuenta que el reconocimiento por voz difiere de la llamada tecnología del reconocimiento de discurso, que reconoce las palabras a medida que el individuo habla y que no es una tecnología biométrica.

Mientras que, el proceso de la identificación de personas a través del reconocimiento de voz depende de diversas características del individuo: por un lado, está la estructura física del tracto vocal, por otro se encuentran ciertas características de comportamiento, En el momento del proceso de identificación se ha de tener muy en cuenta la variabilidad que posee la señal de voz, pues el individuo no puede repetir de forma completamente exacta una misma palabra o frase. Existen dos formas principales de realizar el reconocimiento por voz, dependiendo de si el sistema es dependiente de un texto (una contraseña o una fase programada dentro del sistema), o es independiente de este.

Al respecto, de las muestras de voz obtenidas se analiza el contenido de la frecuencia del discurso, comparando las características de calidad, duración, dinámica, intensidad y tono de la señal. Tras el procesado de la muestra obtenida se realiza la comparación con aquellas almacenadas en la base de datos. En este caso el sistema determinará un ratio de similitud, debido a la variabilidad que caracteriza a la señal de voz.

### **2.2.3 Historia del reconocedor de voz.**

La historia de los primeros Reconocedores de voz comienza hace 40 años, pero los sucesos más importantes que se desarrollaron dentro de la industria fueron gracias a que AT&T introdujo su Sistema de reconocimiento de voz en 1992 para llamadas por cobrar, para a finales de 1993 procesaba 50 millones de llamadas al mes. En 1995 los teléfonos celulares ya ofrecían servicios de marcado activado por voz.

En los 1990's en el sector Salud se inicia la generación de reportes clínicos por voz en salas de emergencia (fomentado por las empresas de seguros Blue Cross y Blue Shield). Posteriormente se amplió a diversas áreas de hospitales como radiología, cardiología, etc.

En 1995 se desarrollan los primeros prototipos de aparatos electrodomésticos Whirlpool Corp.

Actualmente las PC's integran utilerías de voz. Microsoft Corp.: Facilidades para construir objetos de comandos de voz (voice-commandobjects) en Win95. CreativeLabs integra Procesamiento automático en la mayoría de sus tarjetas soundblaster. Compaq y PureSpeech desarrollan conjuntamente tecnología de voz. SeagateTech compró 25% de DragonSystems.

En orden cronológico de los reconocedores de voz:

- 1870's Alexander Graham Bell:

Quería construir un sistema/dispositivo que hiciera el habla visible a las personas con problemas auditivos. Resultado: el teléfono.

- 1880's Tihamir Nemes:

Solicita permiso para una patente para desarrollar un sistema de transcripción automática que identificara secuencias de sonidos y los imprimiera (texto). Pero fue rechazado como "Proyecto no Realista".

- 30 años después AT&T Bell Laboratories:

Construye la primera máquina capaz de reconocer voz (basada en Templates) de los 10 dígitos del inglés. Requería extenso reajuste a la voz de una persona, pero una vez logrado tenía un 99% de certeza. Por lo tanto, surge la esperanza de que el reconocimiento de voz es simple y directo.

- A mediados de los 60's:

La mayoría de los investigadores reconoce que era un proceso mucho más intrincado y sutil de lo que habían anticipado. Por lo tanto, empiezan a reducir los alcances y se enfocan a sistemas más específicos:

- Dependientes del Locutor.

- Flujo discreto de habla (con espacios / pausas entre palabras)

- Vocabulario pequeño (menor o igual a 50 palabras)

Estos sistemas empiezan a incorporar técnicas de normalización del tiempo (minimizar diferencia en velocidad del habla). Además, ya no buscaban una exactitud perfecta en el reconocimiento.

- Principios 1970's:

Se produce el 1er Producto de reconocimiento de voz, el VIP100 de Threshold Technology Inc. (utilizaba un vocabulario pequeño, dependiente del locutor, y reconocía palabras discretas). Gana el U.S. National Award en 1972. Luego: Nace el interés de ARPA del U.S. Department of Defense, y gracias al lanzamiento de grandes proyectos de investigación y financiamiento por parte del gobierno se precipita la época de la inteligencia artificial. El proyecto financiado por ARPA busca el reconocimiento de habla continua, de vocabulario grande. Impulsa que los investigadores se enfoquen al entendimiento del habla. Los sistemas empiezan a incorporar módulos de:

- Ü Análisis léxico (conocimiento léxico)
- Ü Análisis sintáctico (Estructura de Palabras)
- Ü Análisis semántico (Significado)
- Ü Análisis pragmático (Intención)

- 80's a 90's:

Surgen los sistemas de vocabulario amplio, que ahora son la norma. (Más de 1000 palabras). Adicionalmente bajan los precios de estos sistemas. Empresas importantes actualmente:

- Ü Philips
- Ü Lernout&Hauspie
- Ü SensoryCircuits
- Ü DragonSystems
- Ü Speechworks
- Ü Vocalis
- Ü Dialogic
- Ü Novell
- Ü Microsoft
- Ü NEC, Siemens, Intel (apoyo / soporte técnico), entre otros.

#### **2.2.4 Etapas de un sistema de identidad biométrica.**

Las técnicas de identificación biométrica son muy diversas, ya que cualquier elemento significativo de una persona es potencialmente utilizable como elemento de identificación biométrica. Sin embargo, incluso con la diversidad de técnicas existentes, a la hora de desarrollar un sistema de identificación biométrica, se mantiene un esquema totalmente independiente de la técnica empleada.

- **Reclutamiento:**

En esta fase, se toma una serie de muestras del usuario, y se procesan, para posteriormente extraer un patrón, el cual se almacenará y será el conjunto de datos que caracterizará a ese usuario. Si se captura más de una muestra, el patrón suele ser el

resultado de una media de las características obtenidas. Este proceso se hace de forma supervisada, es decir, existe una persona encargada de controlar cómo se produce la captura de los datos, así como de asegurar la identidad de la persona que se está reclutando en el sistema.

- **Utilización:**

Una vez que se tiene almacenado el patrón del usuario, éste puede utilizar el sistema con normalidad, y sus características son comparadas con el patrón almacenado, determinando el éxito o fracaso de esa comparación. Cada una de las fases mencionadas, está basada en una serie de bloques que hacen que las características biológicas o de comportamiento del individuo acaben siendo un elemento que lo identifique.

- **Captura:**

Se toman los datos biofísicos o de comportamiento del sujeto. La toma de los datos depende, evidentemente, de la técnica biométrica empleada, también se pueden encontrar muchas variaciones una misma técnica biométrica. Por ejemplo, la huella dactilar puede ser obtenida por cámara de video, ultrasonidos, efecto capacitivo sobre un semiconductor o exploración por láser. Esta fase es muy importante ya que en ella se está contenida la interfaz hombre-máquina y el sensor para la captura de la información biométrica, esto se repercute directamente en el rendimiento del sistema biométrico ya que un diseño pobre de la interfaz puede resultar en una tasa alta de fallos al adquirir la información.

- **Comparación:**

Una vez extraídas las características de la muestra capturada, se han de comparar estas con las previamente almacenadas, es decir, el patrón o plantilla. Lo más importante que hay que dejar claro cuando se habla de este bloque, es que no se trata de una comparación binaria (o de igualdad), sino que la variación de las muestras, por diferencias en la captura o leve variación de las características de sujeto, hacen que la comparación de como resultado un puntaje o probabilidad de semejanza. Por tanto para

determinar el éxito o fracaso de la comparación, habrá que determinar un umbral de tolerancia en esa probabilidad.

- **Reconocimiento:**

También llamado, en algunos textos, simplemente identificación, se basa en identificar a un usuario dentro de todos los usuarios que ya se encuentran en el sistema. Por lo tanto, se comparan las características extraídas con los patrones de todos los usuarios reclutados por el sistema. Este esquema de funcionamiento, necesario para muchas aplicaciones, tiene como inconvenientes la necesidad de una base de datos de patrones y la existencia de una red de comunicaciones, siempre online, que comunique los puestos de identificación la base de datos. El resultado de la comparación puede ser: siempre positivo, o puede indicar rechazos.

- **Autenticación:**

También llamado sencillamente verificación. Trata de responder a la pregunta ¿es este sujeto la persona que dice ser? En este esquema de funcionamiento, el usuario, al que se le toman sus características biométricas, también comunica su

- **Identidad:**

El sistema se encarga, entonces, de comparar las características extraídas, con el patrón del usuario indicado. Si la comparación supera un determinado umbral de similitud, se considera que el usuario es el indicado, rechazando la comparación en caso contrario. El patrón del usuario puede estar almacenado en una base de datos, tal y como se hace en los sistemas de reconocimiento, o, si el patrón es suficientemente pequeño, en un sistema portátil de información como puede ser una tarjeta. En este último caso no son necesarias ni la base de datos ni la red de comunicación de los sistemas de reconocimiento.

### **2.2.5 Características de la señal de voz:**

Los sistemas de reconocimiento de voz se enfocan en las palabras y los sonidos que distinguen una palabra de la otra en un idioma. Estas son los fonemas. Por ejemplo,

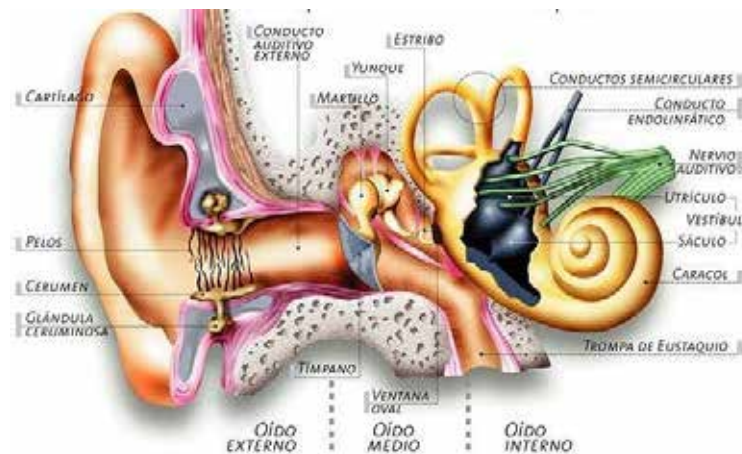
"tapa", "capa", "mapa", "napa", son palabras diferentes puesto que su sonido inicial se reconoce como fonemas diferentes en español.

Existen varias maneras para analizar y describir el habla. Los enfoques más comúnmente usados son:

- **Articulación:** Análisis de cómo el humano produce los sonidos del habla.
- **Señal Acústica:** Análisis de la señal de voz como una secuencia de sonidos.
- **Percepción Auditiva:** Análisis de cómo el humano procesa el habla.

Los tres enfoques proveen ideas y herramientas para obtener mejores y más eficientes resultados en el reconocimiento.

En la figura 1, se observa que partes del oído intervienen en el proceso de comunicación del habla.



**Figura 1.**Proceso de comunicación del habla

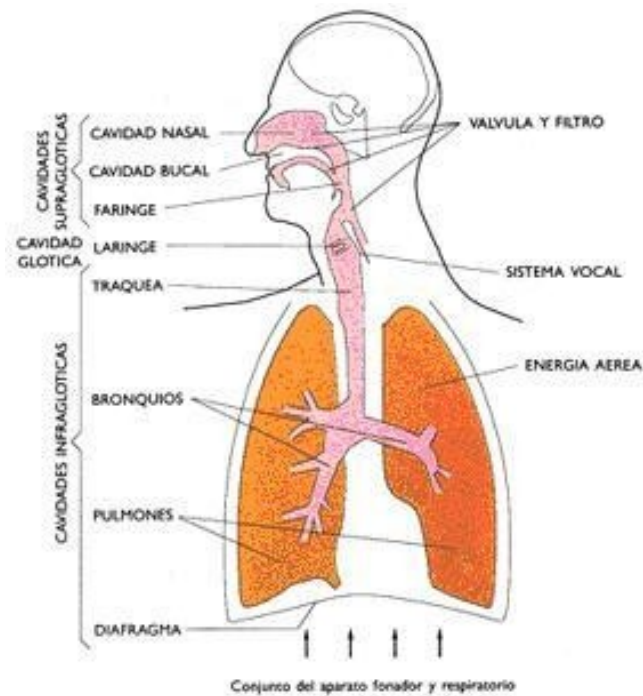
Fuente: Castaños, E. Percepción del sonido: la audición.

· **Articulación:**

La Articulación se compone del sistema fonador que es un conjunto de órganos que intervienen en la producción de sonidos. También llamado aparato bucal o articulario.

El aparato fonador consiste en tres grupos de órganos diferenciados (ver Figura 2):

- Órganos de respiración-cavidades infra-glóticas: pulmones, bronquios y tráquea.
- Órganos de fonación-cavidades glóticas: laringe, cuerdas vocales y resonadores - nasal, bucal y faríngeo.
- Órganos de articulación-cavidad supra-glóticas: paladar, lengua, dientes, labios y glotis.



**Figura 2.** Órganos del cuerpo que intervienen en el aparato fonador.  
Fuente: Taller de técnica vocal EMMAT.

El correcto funcionamiento del aparato fonador, es controlado por el sistema nervioso central. Específicamente, se sabe que el control del habla se realiza en el área de Broca que se encuentra en el hemisferio izquierdo de la corteza cerebral.

La articulación centra su atención en el aparato vocal: Garganta, boca, nariz, en donde se producen los sonidos del habla.

- Garganta: Contiene las cuerdas vocales, cuya vibración produce los fonemas "hablados".

- Boca y Nariz: "Cavidades de resonancia" por que refuerzan ciertas frecuencias sonoras.
  - Ü Cuando el paladar suave baja y deja el aire pasar por la nariz se generan los fonemas nasales (/m/ /n/)
  - Ü La boca consiste de:
    - Puntos de articulación.
    - Dientes.
    - Puente alveolar (puente óseo atrás de los dientes superiores).
    - Paladar duro.
    - Paladar suave o velum.
    - Articuladores.
    - Labios.
    - Lengua.

### **2.2.6 Señal acústica.**

Un reconecedor no puede analizar los movimientos en la boca. En su lugar, la fuente de información es la señal de voz misma.

El Habla es una señal analógica, es decir, un flujo continuo de ondas sonoras y silencios.

El conocimiento de la ciencia de la acústica se utiliza para identificar y describir los atributos del habla que son necesarios para un reconocimiento de voz efectivo.

Algunas características importantes del análisis acústico son:

- **Frecuencia y amplitud.**

Todos los sonidos causan movimientos entre las moléculas del aire. Algunos sonidos, tales como los que produce una cuerda de guitarra, producen patrones regulares y prolongados de movimiento del aire. Los patrones de sonidos más simples son los sonidos puros (pure tones), y se pueden representar gráficamente por una onda sinusoidal.

Frecuencia: Número de vibraciones (ciclos) del tono por segundo 100 por segundo = 100 ciclos/segundo = 100 Hz.

Tonos altos = Mayor frecuencia

Tonos bajos = Menor frecuencia

El volumen de un sonido refleja la cantidad de aire que es forzada a moverse. Se describe y representa como amplitud de la onda y se mide en decibeles dB.

- **Resonancia.**

La resonancia se define comúnmente como la habilidad que tiene una fuente vibrante de sonido de causar que otro objeto vibre gracias a ella.

La mayoría de los sonidos incluyendo del habla tienen una frecuencia dominante llamada frecuencia fundamental también conocida como pitch (tono) que se combina con frecuencias secundarias en el habla, la frecuencia fundamental es la velocidad a la que vibran las cuerdas vocales al producir un fonema sonoro.

Sumadas a la frecuencia fundamental hay otras frecuencias que contribuyen al timbre del sonido. (Son las que nos permiten distinguir una trompeta de un violín, etc. o las voces de diferentes personas)

Algunas bandas de la frecuencia secundarias juegan un rol importante en la distinción de un fonema de otro. Se les llama formantes y son producidas por la resonancia.

La garganta, boca y nariz son cámaras de resonancia que amplifican las bandas o frecuencias formantes contenidas en el sonido generado por las cuerdas vocales. Estas formantes amplificadas dependen del tamaño y forma de la boca y si el aire pasa o no por la nariz.

Los patrones de las formantes son más fuertes (distinguibiles) para vocales que para las consonantes no sonoras.

- **Estructura armónica y ruido.**

El habla no es un tono puro es continuación de múltiples frecuencias y se representa como una onda compleja.

Vocales se componen de 2 o más ondas simples son ricos en frecuencias secundarias y contienen estructuras internas que incluyen ondas cíclicas y a cíclicas.

Las ondas a cíclicas no tienen patrones repetitivos generalmente llamados ruido forman parte de todos los fonemas sonoros, consonantes y semivocales.

Las frecuencias y características de los patrones a cíclicos proveen información importante sobre la identidad de los fonemas.

La identidad de las consonantes también se revela por el cambio en las formantes que resultan cuando los articuladores se mueven de un fonema anterior a la consonante y de ella al siguiente fonema llamadas transiciones de formantes. Estas se analizan utilizando técnicas como la transformada rápida de Fourier (FFT) generando espectrogramas.

La complejidad de las formas de onda de los fonemas y las constantes transiciones de un patrón a otro dificultan el análisis de los patrones utilizando las representaciones complejas de las ondas.

Los patrones armónicos y de ruido se muestran con más claridad utilizando los espectrogramas de banda ancha.

La localización (la distancia entre ellas) y cambio en las formantes ayudan a identificar fonemas y palabras.

- **Coarticulación.**

Los fonemas aparentemente tienen parámetros acústicos claramente definidos, pero más bien, los fonemas tienden a ser abstracciones implícitamente definidas por la pronunciación de palabras en un lenguaje.

La forma acústica de un fonema depende fuertemente del contexto acústico en el que sucede a éste efecto se le llama coarticulación.

Investigadores, utilizan este concepto para distinguir entre la característica conceptual de un sonido del habla (fonema) y una instancia o pronunciación específica de ese fonema (tono).

### **2.2.7 Percepción auditiva.**

La variabilidad del habla producida por coarticulación y otros factores hacen del análisis de la voz extremadamente difícil.

La facilidad del humano en superar estas dificultades sugiere que un sistema basado en la percepción auditiva podría ser un buen enfoque desafortunadamente el conocimiento de la percepción humana es incompleto, lo que se sabe es que el sistema auditivo está adaptado a la percepción de la de voz.

El oído humano detecta frecuencias de 20Hz a 20,000 Hz pero es más sensible al rango entre 1000 y 6000 Hz. También es más sensible a cambios pequeños en la frecuencia en el ancho de banda crítico para el habla. Además el patrón de sensibilidad a cambios en el tono (pitch) no corresponde a la escala lineal de frecuencias de ciclos por segundo de la acústica.

Para representar mejor al patrón de percepción acústica, se tiene una escala llamada mel-scale, la cual es una escala logarítmica que representa los niveles de la señal.

El humano no procesa frecuencias individuales independientemente, como lo sugiere el análisis acústico. En su lugar escucha grupos de frecuencias por lo cual es capaz de distinguirlas de ruidos alrededor.

### **2.2.8 Micrófono.**

Es un transductor electro acústico, que tiene como función transformar o traducir la presión acústica ejercida sobre su capsula por las ondas sonoras en energía eléctrica. La calidad de cada micrófono viene dada por sus características, las cuales se describen a continuación:

- **Sensibilidad**

Es la eficiencia del micrófono, la relación entre la presión sonora que incide (expresada en Pascales) y también la tensión eléctrica de salida (expresada en voltios). Es decir, expresa que tan bien convierte el micrófono la presión acústica en voltaje de salida. La sensibilidad se expresa en milivoltios por Pascal.

Al utilizar el milivoltio, la sensibilidad puede ser representada en un voltímetro de la siguiente manera: a mayor voltaje, mayor sensibilidad.

- **Fidelidad**

Indica la variación de sensibilidad con respecto a la frecuencia. Además, la fidelidad, viene definida como la respuesta en frecuencia del micrófono, cuanto más lineal sea la respuesta en frecuencia mayor fidelidad tendrá el micrófono. La fidelidad se expresa en dB.

En función de esta respuesta en frecuencia o fidelidad se elabora la llamada curva de respuesta de un micrófono, que es la representación gráfica del nivel obtenido en la captación de sonidos de igual intensidad, pero de distinta frecuencia.

- **Directividad**

Esta característica determina en qué dirección capta mejor el sonido un micrófono, es decir indica la sensibilidad del micrófono a las diferentes direcciones. El diagrama polar es una representación gráfica que indica que tan sensitivo es un micrófono a los sonidos que llegan a él desde diferentes ángulos alrededor de su eje central.

Dependiendo de la directividad se encuentran diferentes tipos de micrófonos:

- Û **Omnidireccionales:** Captan los sonidos, sin importar la dirección donde lleguen
- Û **Bidireccionales:** captan tanto el sonido que llega por su parte frontal, como por su parte posterior.
- Û **Unidireccionales o direccionales:** captan el sonido en una sola dirección mientras que son relativamente sordos a las otras direcciones.

- **Ruido de fondo**

Es la tensión que entrega el micrófono sin que exista ningún sonido incidiendo sobre él. Este ruido se produce por el movimiento térmico de los electrones en la carcasa que no tiene masa. El ruido de fondo debe estar en torno a los 60 dB, pero mientras más bajo sea, mejor calidad ofrece el micrófono.

- **Rango dinámico**

Es el margen que hay entre el nivel de referencia de salida máxima y el ruido de fondo de un determinado sistema, medido en decibelios. En este caso, el rango dinámico y relación señal/ruido.

- **Impedancia interna**

Es la resistencia que opone el micrófono al paso de la corriente. La impedancia según su valor viene caracterizada por baja, alta y muy alta impedancia.

- ü Lo-Z Baja impedancia (alrededor de 200 Ohmios).

- ü Hi-Z Alta impedancia (1 o 3 k ohm e incluso 600 ohm).

- ü VHi-Z Muy alta impedancia (más de 3 k ohm).

Por otra parte es importante destacar que si el micrófono es de alta impedancia y se tiene un cable largo se produce una pérdida muy grande. Si se tiene una impedancia baja se puede utilizar un cable muy grande. Si se tiene una impedancia baja se puede utilizar un cable muy largo y no se pierde tanto la señal. Por último, se puede bajar la resistencia para evitar pérdidas en altas frecuencias.

### **2.2.9 Muestreo y Cuantización.**

El habla es una señal continua y varía en el tiempo. Las variaciones en la presión del aire se irradian desde la cabeza y se transmiten por el aire.

Un micrófono convierte esas variaciones en presión del aire a variaciones en voltaje. Una señal analógica se puede transmitir a través de un circuito telefónico (voltaje) o almacenados en una cinta magnética (flujo magnético).

En el mundo real los estímulos sensoriales son análogos. Sin embargo, para las computadoras es necesario digitalizar la señal.

Para esto se usan una serie de valores numéricos con una frecuencia regular (frecuencia de muestreo). El número posible de valores está limitado por el número de bits disponible para representar a cada muestra.

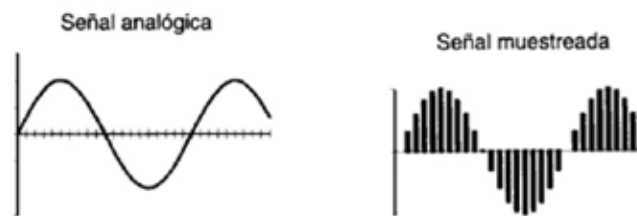
- **Muestreo**

El muestreo asigna un valor numérico a la señal en unidades discretas de tiempo constante dependiendo de la frecuencia de Nyquist que especifica la frecuencia máxima a la que una señal puede reproducirse completamente.

El teorema de Nyquist establece que:

Frecuencia de muestreo=2 veces la frecuencia máxima de la señal para poder reproducirla.

Para poder reproducir la señal analógica debe pasar por un filtro pasa-bajas a la frecuencia de muestreo (quitar ruidos creados por el muestreo)



**Figura 3.** Señal senoidal muestreada.

Fuente: Kuhlmann. F y Alonso. A Señales, sistemas y sus características

Si la señal varía más rápido se requiere una  $T$  más pequeña por lo tanto un menor ancho de banda de frecuencias. El humano produce señales de voz desde los 100(hombre)-400(mujer) Hz hasta los 15000 Hz.

#### · **Cuantización**

La cuantización consiste en que cada muestra se representa con un valor digital limitando el rango de valores discretos correspondiente a la original.

Esto se puede ver como la resolución:

- ü El error o diferencia entre la señal original y la reconstrucción se percibe como ruido.
- ü Por lo tanto, a mayor resolución mayor cuantización y menor ruido como consecuencia.
- ü La resolución se describe generalmente en términos de la relación señal a ruido.

Ü A mayor SRN es mayor la fidelidad de la señal digitalizada.

Ü SNR aproximadamente  $2^B$  ( B=bits/muestra)

Ü Es independiente de la frecuencia de muestreo.

Existen diferentes técnicas de cuantización:

- **Cuantización uniforme**

En los cuantificadores uniformes o lineales la distancia entre los niveles de reconstrucción es siempre la misma, la mayoría usan un número de niveles que es una potencia de 2. No hacen ninguna suposición acerca de la señal a cuantificar, de allí que no proporcionen los mejores resultados.

Su ventaja es que son más fáciles y económicos al implementarlos

- **Cuantización no uniforme**

El problema de la cuantificación uniforme es que conforme aumenta la amplitud de la señal, también aumenta el error. Este problema lo resuelve el cuantificador logarítmico de forma parcial. Sin embargo, si conocemos la función de la distribución de probabilidad, podemos ajustar los niveles de reconstrucción a la distribución de forma que se minimice el error cuadrático medio. Esto significa que la mayoría de los niveles de reconstrucción se den en la vecindad de las entradas más frecuentes y, consecuentemente, se minimice el error (ruido).

- **Cuantización logarítmica**

Las señales de voz pueden tener un rango dinámico superior a los 60 dB, por lo que para conseguir una alta calidad de voz se deben usar un elevado número de niveles de reconstrucción. Sin embargo, interesa que la resolución del cuantificador sea mayor en las partes de la señal de menor amplitud que en las de mayor amplitud. Por tanto, en la cuantificación lineal se desperdician niveles de reconstrucción y, consecuentemente, ancho de banda. Esto se puede mejorar incrementando la distancia entre los niveles de reconstrucción conforme aumenta la amplitud de la señal.

Para evitar desperdicio de niveles de reconstrucción y ancho de banda se utiliza un método sencillo para mejorar el incremento de la distancia entre los niveles de

reconstrucción conforme aumenta la amplitud de la señal. Para conseguir esto se hace pasar la señal por un compresor logaritmo antes de la cuantificación. Esta señal comprimida puede ser cuantificada uniformemente. A la salida del sistema de la señal pasa por un expansor. A esta técnica se le llama compresión.

- **Cuantización vectorial**

Este método cuantifica los datos en bloques de N muestras. En este tipo de cuantificación, el bloque de N muestras se trata como un vector N-dimensional. La cuantificación vectorial ofrece mejores resultados que la cuantificación escalar, sin embargo, es más sensible a los errores de transmisión y lleva consigo una mayor complejidad computacional.

### **2.2.10 MATLAB.**

Matlab es el nombre abreviado de “MATrix LABoratory”. Es un lenguaje de alto nivel y de ambiente interactivo que permite realizar tareas intensas y con una mayor velocidad que los lenguajes de programación comúnmente usados.

Matlab se especializa en cálculos numéricos con vectores y matrices, como casos particulares puede trabajar también con otras estructuras de información. Aunque cada objeto es considerado como un arreglo.

El lenguaje está construido por un código llamado M-code que puede ser fácilmente ejecutado en la ventana de comandos. Con lo cual pueden crear funciones, etc. Para la razón principal para la elección de este lenguaje de programación son las herramientas que proporcionan para el procesamiento de señales, y el conjunto de funciones para el procesamiento digital.

Además, para crear entornos gráficos se pueden utilizar el GUIDE de Matlab, que provee herramientas para crear GUI's, “Graphical User Interface”, con lo cual se puede crear la forma del entorno grafico así como asociar funciones a los elementos del GUI. Matlab también incluye funciones para manipular diversos archivos.

### **2.2.11 Obtención de información mediante el micrófono.**

Para grabar una señal de voz se hace la conversión de la señal analógica del micrófono en una señal digital por medio del conversor A/D en la tarjeta de sonido. Cuando un micrófono está operando las ondas de sonido hacen que vibre el elemento magnético del micrófono causando una corriente eléctrica hacia la tarjeta de sonido, donde el conversor A/D básicamente graba los voltajes eléctricos en intervalos específicos.

Hay dos factores importantes durante este proceso. Primero está la tasa de muestreo o que tan seguido los valores de voltaje son grabados. Segundo, son los bits por segundo, o que tan exactamente los valores son grabados. Un tercero podría ser el número de canales (mono o estéreo), pero para las mayoría de aplicaciones vienen con valores predeterminados, para desarrollo del código se debería cambiar los parámetros para ver lo que mejor funciona en el algoritmo.

Dado que el habla es relativamente de bajas frecuencias (entre 100 Hz y 8KHz), una frecuencia de muestreo de 16000 muestras/seg provee una mayor exactitud en la adquisición de la información, la frecuencia de Nyquist.

Es importante destacar que la obtención de la información mediante micrófono en Matlab se realiza con la función wavrecord (n, fs) graba n muestras de la señal de audio, muestreadas a una cierta frecuencia.

### **2.2.11 Métodos de reconocimiento de voz.**

En el reconocimiento del habla, la señal de voz ingresa a un procesamiento para producir una representación de la voz en forma de secuencia de vectores o agrupaciones de valores que se denominan parámetros, que deben representar la información contenida en la envolvente del espectro.

Es importante destacar que el número de parámetros debe ser reducido, para no saturar la base de datos ya que mientras más parámetros tengan la representación menos fiables son los resultados y más costosa es la implementación.

Existen distintos métodos de análisis para la extracción de características, y se concentran en diferentes aspectos representativos, las dos de mayor importancia para el análisis de la voz son las siguientes:

ü Análisis de predicción lineal (LPC)

ü Análisis Cepstral

· **Predicción lineal.**

Es uno de los métodos más útiles para codificar la voz con una buena calidad, su función es representar la envolvente espectral de una señal digital de voz en una forma comprimida, utilizando la información de un modelo lineal, con lo cual se proporcionan unas aproximaciones a los parámetros de la voz muy precisas.

Este método pretende extrapolar el valor de la siguiente muestra de voz  $s(n)$  como la suma ponderada de muestras pasadas:

$$s(n) = s(n$$

espectro nos proporciona información acerca de los parámetros del modelo de producción de voz.

Estos coeficientes serán utilizados solo para mostrar la señal acústica más no para realizar ningún tipo de cálculo, estimaciones, entre otros.

### **2.3 Definición de términos básicos.**

**Algoritmo:** Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema.

**Amplitud:** Distancia entre los puntos más alto y más bajo de la onda, determina la intensidad. Es una medida de la variación máxima del desplazamiento u otra magnitud física que varía periódica o cuasi-periódicamente en el tiempo.

**Armónico:** En una onda periódica, cualquiera de sus componentes sinusoidales, cuya frecuencia sea un múltiplo entero de la frecuencia fundamental.

**Biometría:** es el estudio para el reconocimiento inequívoco de personas basadas en uno o más rasgos conductuales.

**Cuantización:** consiste en que cada muestra se representa con un valor digital limitando el rango de valores discretos correspondiente a la original.

**Distorsión:** deformación de un sonido, una imagen, una señal, etc., producida durante su transmisión o reproducción.

**Espectro:** Distribución de la intensidad de una radiación en función de una magnitud característica, como la longitud de onda, la energía, la frecuencia o la masa.

**Filtro:** Dispositivo que elimina o selecciona ciertas frecuencias de un espectro eléctrico, acústico, óptico o mecánico, como las vibraciones.

**Fonemas:** Sonido de la voz

**Frecuencia:** Número de veces que se repite un proceso periódico por unidad de tiempo.

**Impedancia:** Relación entre la tensión alterna aplicada a un circuito y la intensidad de la corriente producida. Se mide en ohm.

**Interfaz:** Conexión física y funcional entre dos aparatos o sistemas independientes.

**Muestreo:** asigna un valor numérico a la señal en unidades discretas de tiempo constante dependiendo de la frecuencia Nyquist que especifica la frecuencia máxima a la que una señal puede reproducirse completamente.

**Nítido:** Limpio, terso, claro, puro.

**Normalizar:** Es el proceso de elaborar, aplicar y mejorar las normas que se aplican a distintas actividades científicas, industriales o económicas con el fin de ordenarlas y mejorarlas. La normalización persigue fundamentalmente tres objetivos:

- Simplificación: se trata de reducir los modelos para quedarse únicamente con los más necesarios.
- Unificación: para permitir el intercambio a nivel internacional.
- Especificación: se persigue evitar errores de identificación creando un lenguaje claro y preciso.

**Pronunciar:** Emitir y articular sonidos para hablar.

**Resonancia:** sonido elemental que, junto con otros, acompaña al principal en una nota musical y comunica un timbre particular a cada voz o instrumento.

**Ruido:** Segmento: Porción o parte cortada o separada de una cosa, de un elemento geométrico o de un todo.

**Sensibilidad:** Capacidad de respuesta a muy pequeñas excitaciones, estímulos o causas.

**Silencio:** El silencio surge cuando no hay sonidos perceptibles, pero en realidad es un estado mental, porque siempre suena algo, aunque sea muy levemente.

**Sistemas biométricos:** un sistema biométrico es básicamente un sistema de reconocimiento de patrones que reconoce a un individuo estableciendo la autenticidad de la biometría que posee.

**Sonido:** Sensación producida en el órgano del oído por el movimiento vibratorio de los cuerpos, transmitido por un medio elástico, como el aire.

**Timbre:** Calidad de los sonidos, que diferencia a los del mismo tono y depende de la forma y naturaleza de los elementos que entran en vibración.

**Tono:** Cualidad de los sonidos, dependiente de su frecuencia, que permite ordenarlos de graves a agudos.

**Transductor:** Dispositivo que transforma el efecto de una causa física, como la presión, la temperatura, la dilatación, la humedad, etc., en otro tipo de señal, normalmente eléctrica.

**Ventanas de Haming:** En procesamiento de señales, una ventana de Haming es un zoom de una señal de longitud voluntariamente limitada. Sonido inarticulado, por lo general desagradable.

**Voz:** sonido producido por el ser humano o los animales cuando el aire es expulsado a través de la laringe y hace que vibren las cuerdas vocales.

## **CAPÍTULO III MARCO METODOLÓGICO**

Marco metodológico según Arias (2006): “La metodología de proyecto incluye el tipo o tipos de investigación, las técnicas y los instrumentos que serán utilizados para llevar a cabo la indagación. Es el “como” se realizará el estudio para responder al problema planteado.

### **3.1 Tipo de investigación.**

Con respecto al tipo de investigación, Tamayo, M (2003) expresa sobre la investigación descriptiva “Comprende la descripción, registro, análisis e interpretación de la naturaleza actual, y la composición o procesos de los fenómenos. El enfoque se hace sobre conclusiones dominantes o sobre como una persona, grupo o cosa se conduce o funciona en el presente. La investigación descriptiva trabaja sobre realidades de hecho, y su característica fundamental es la de presentarnos una interpretación correcta.”

En relación con lo expresado anteriormente, se dice que la presente investigación puede calificarse como documental – descriptiva, pues constituye un estudio sistemático de investigaciones previas ya comprobadas, y a su vez, se realiza bajo el esquema de un proyecto factible, cuyo enfoque se centra en la posibilidad de llevar teorías generales al ámbito práctico, y cuyo esfuerzo se destina a la implantación de propuestas, que pueden materializarse y brindar soluciones a problemas que se plantean en la sociedad.

### **3.2 Diseño de la investigación.**

El diseño de la investigación es el conjunto de directrices que toma el investigador con el fin de observar, analizar y plantear una solución de ser posible a la problemática objeto de la investigación. Según Palella, S y Martins, F (2012) definen como investigación de campo a:

“La Investigación de campo consiste en la recolección de datos directamente de la realidad donde ocurren los hechos, sin manipular o controlar las variables. Estudia los fenómenos sociales en su ambiente natural. El investigador no manipula variables debido a que esto hace perder el ambiente de naturalidad en el cual se manifiesta”.

### **3.3 Nivel de la investigación.**

El nivel de investigación se refiere según Arias, F (2012) “al grado de profundidad con que se aborda un objeto o fenómeno”. Así pues, el nivel de investigación establece hasta qué punto se llevará a cabo el estudio del tema o problema planteado. Tomando en cuenta el tipo de investigación, se conocerá el nivel en el cual se basa todo el estudio. También el nivel permite saber qué factores tienen que intervenir para el desarrollo de toda la investigación.

Tomando en cuenta lo anteriormente expuesto, el nivel de investigación que se emplea es proyectivo definido por Hurtado, J (2010), como: “es el que intenta proponer soluciones a una situación determinada. Implica explorar, describir, explicar y proponer alternativas de cambio y no necesariamente ejecutar la propuesta”.

Lo mencionado por Hurtado, se aplica a todas las investigaciones que conllevan a diseños dirigidos a cubrir necesidades y que están basados en conocimientos anteriores.

### **3.4 Técnicas e instrumentos de recolección de datos**

#### **3.4.1 Técnicas de recolección de datos**

Según Arias (1999, p.25) las técnicas de recolección de datos son:

Las distintas formas o maneras de obtener la información. Son ejemplos de técnicas; la observación directa, la encuesta en sus dos modalidades (entrevista o cuestionario), el análisis documental, el análisis de contenido, etc.

Las técnicas utilizadas en el proyecto de investigación son las siguientes:

- **Observación directa**

Wilson (2000) comento lo siguiente:

“Es una técnica que consiste en observar atentamente el fenómeno, hecho o caso, tomar información y registrarla para su posterior análisis. La observación es un elemento fundamental de todo proceso investigativo; en ella se apoya el investigador para obtener el mayor número de datos”.

Teniendo en cuenta lo anterior, la observación directa se utilizó para recabar datos de la población sobre la cual se hizo el estudio, para luego pasar a analizar dichos datos.

### **3.4.2 Instrumentos de recolección de datos.**

Arias (1999, p.25) se refirió a los instrumentos de recolección de datos como:

“Los instrumentos son los medios materiales que se emplean para recoger y almacenar la información. Ejemplo: fichas, formatos de cuestionarios, guías de entrevistas, guías de cotejo, grabadoras, escalas de actitudes u opinión.

- **Anotaciones de la observación directa**

Hernández et al. (2010), se refirió a las anotaciones de la observación directa como:

“Descripciones de lo que estamos viendo, escuchando, olfateando y palpando del contexto y de los casos o participantes observados. Regularmente van ordenadas de manera cronológica. Nos permitirán contar con una narración de los hechos ocurridos (qué, quién, cómo, cuándo y dónde).”

Tomando como base la definición anterior, las anotaciones de la observación directa se utilizaron para llevar un control esquemático sobre todos los datos extraídos de la observación directa de la población a la cual se sometió el estudio.

### **3.5 Fases de la investigación.**

**Fase I: Caracterizar la voz humana mediante métodos acústicos, matemáticos y físicos.** Se estudió la voz humana, pudiendo así conocer cuantitativamente las características que posee la voz humana.

**Fase II: Clasificar los datos obtenidos para poder crear una base de datos y utilizarlos en el sistema de seguridad.** Se resumieron los resultados obtenidos, para así crear una base de datos funcional en el sistema de seguridad.

**Fase III: Desarrollar un programa con lenguaje de alto nivel que funcione como interfaz de usuario para el reconocimiento de la voz.** Se programó en MATLAB estableciendo las variables en estudio referentes al reconocimiento de voz, también se creó la interfaz que funciona como mediador con el usuario y el programa.

**Fase IV: Evaluar el programa de reconocimiento de voz mediante la simulación.** En esta etapa, el sistema se simuló y validó el reconocimiento del usuario, se indicó en la interfaz gráfica de que usuario corresponde la voz y le permitió el acceso. Al no reconocer, la acción que realizó es no dar acceso al sistema.

## CAPÍTULO IV RESULTADOS

### 4.1 Fase I: Caracterizar la voz humana mediante métodos acústicos, matemáticos y físicos.

En función de la teoría desarrollada en el Capítulo II se puede decir que la voz para poder ser estudiada tiene que pasar por un proceso por medio del cual se podrá obtener a su salida una señal digital la cual podrá ser observada, estudiada y visualizada de una manera más eficiente. A continuación, será descrito el proceso por el cual va pasar la voz para poder ser estudiada.

#### Generación de la señal acústica

En la figura 4 se puede apreciar lo que conformaría un sistema por medio del cual se genera la voz, es una forma de describir como la misma se produce y se irradia por el aire.



**Figura 4.**Diagrama de generación de la señal acústica  
**Fuente:** Izquierdo, M y Maldonado, A (2019).

### Conversión A/D o adquisición

En la figura 5 se puede apreciar el proceso por medio del cual tiene que pasar la voz para lograr obtener señal eléctrica la cual podrá ser estudiada de mejor forma, este proceso será explicando mejor mas adelante.



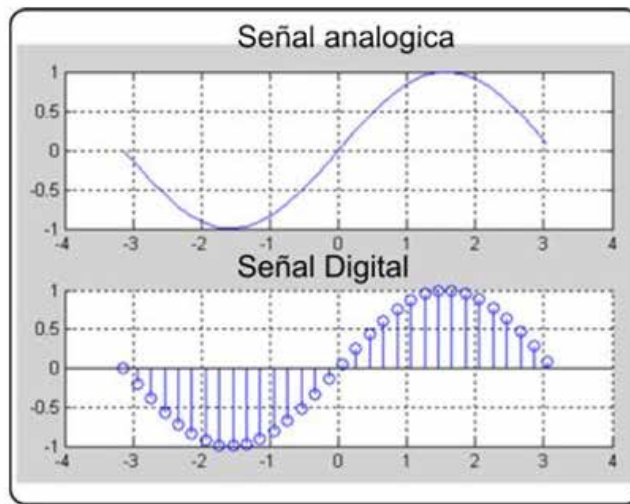
**Figura 5.** Diagrama de la adquisición de voz  
Fuente: Izquierdo, M y Maldonado, A (2019)

### Obtención de la señal por el micrófono

Cuando el micrófono está operando las ondas de sonido hacen que vibre el elemento magnético del micrófono causando una corriente eléctrica hacia la tarjeta de sonido, donde el convertidor A/D graba los voltajes eléctricos en intervalos específicos.

### Muestreo

El muestreo es utilizado para transformar una señal analógica o señal continua a una señal digital o discreta, es decir la señal original la cual es analógica se procede a tomar muestras cada terminado tiempo y se hace una copia de su valor en tiempos definidos, de esta forma surge la frecuencia de muestreo que es la cantidad de muestras que se tienen de una señal en una unidad de tiempo y se mide en Hz los cuales son ciclos por segundo.



**Figura 6.** Señal tratada digitalmente para ser muestreada  
**Fuente:** Izquierdo, M y Maldonado, A (2019)

### **Cuantificación**

El proceso de cuantificación es uno de los pasos que se siguen para lograr la digitalización de una señal analógica. Este proceso es convertir una sucesión de muestras de amplitud continua en una sucesión de valores discretos preestablecidos según el código utilizado.

Durante el proceso de cuantificación se mide el nivel de tensión de cada una de las muestras, obtenidas en el proceso de muestreo, y se les atribuye un valor finito de amplitud, seleccionado por aproximación dentro de un margen de niveles previamente fijado.

### **Ganancia**

A continuación, se amplifican las señales a niveles que sean manejables. La señal vocal tiene componentes de frecuencia que pueden llegar a los 10kHz, sin embargo, la mayor parte de los sonidos vocales tienen energía espectral significativa de hasta los 5kHz.

## Pre-procesamiento

Para realizar el análisis de la señal de voz se debe realizar un pre-procesamiento de la señal vocal. Esto sucede a través de técnicas que permitan extraer la información acústica directamente a partir de la señal vocal emitida, se realiza el muestreo y la cuantización mencionados anteriormente.



**Figura 7.** Diagrama de la digitalización de la señal de voz

**Fuente:** Izquierdo, M y Maldonado, A (2019)

## Eliminación de ruido y normalización

Se realiza un filtrado de la señal para reducir los efectos del ruido que acompaña la señal de voz ingresada, sin este filtrado se tendrían datos erróneos o innecesarios.

Por otra parte, también se realiza una segmentación, esto permite detectar el inicio y el final de la señal de voz para eliminar el ruido inicial y final, lo cual permite optimizar memoria y hacer más veloz el tiempo de cálculo para comparaciones y por supuesto evitar errores, para hacer esto se hace uso del nivel de energía.

## Filtro pre-énfasis

Se aplica un filtro digital pasa altas de primer orden a la señal, para enfatizar las frecuencias altas de los formantes por dos razones, primero para que no se pierda información durante la segmentación, ya que la mayoría de la información está contenida en las frecuencias bajas, en segundo remueve la componente DC de la señal, aplanando espectralmente la señal.

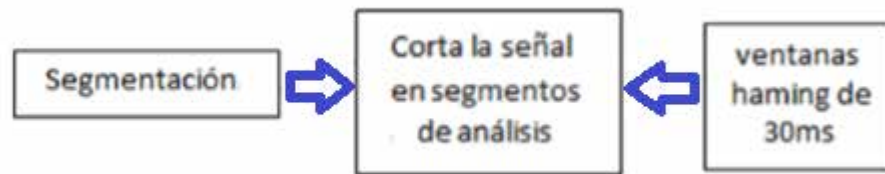


**Figura 8.** Diagrama de los filtros de énfasis  
Fuente: Izquierdo, M y Maldonado, A (2019)

## Segmentación

La segmentación consiste en cortar la señal en segmentos de análisis. La señal de voz es asumida como estacionaria en estos segmentos.

Durante la segmentación los segmentos son guardados cada uno como la columna de una matriz, para el posterior procesamiento de la señal de voz.



**Figura 9.** Diagrama de segmentación, ventaneo y recorte de la señal acústica  
**Fuente:** Izquierdo, M y Maldonado, A (2019)

### Extracción de las características

En este paso se utiliza el método de predicción lineal el cual fue mencionado en el Capítulo II para obtener los coeficientes de LPC y con ayuda de Cepstrum entender el espectro de frecuencia de la señal que se va a reconocer.

Por otra parte, es de hacer notar que los coeficientes de cepstrales tienen la ventaja de que a partir de ellos se puede derivar una serie de parámetros que son invariantes sin importar las distorsiones que puedan ser introducidas.



**Figura 10.** Diagrama que explica los procesos de extracción de características  
**Fuente:** Izquierdo, M y Maldonado, A (2019)

### 4.2 Fase II: Clasificar los datos obtenidos para poder crear una base de datos y utilizarlos en el sistema de seguridad.

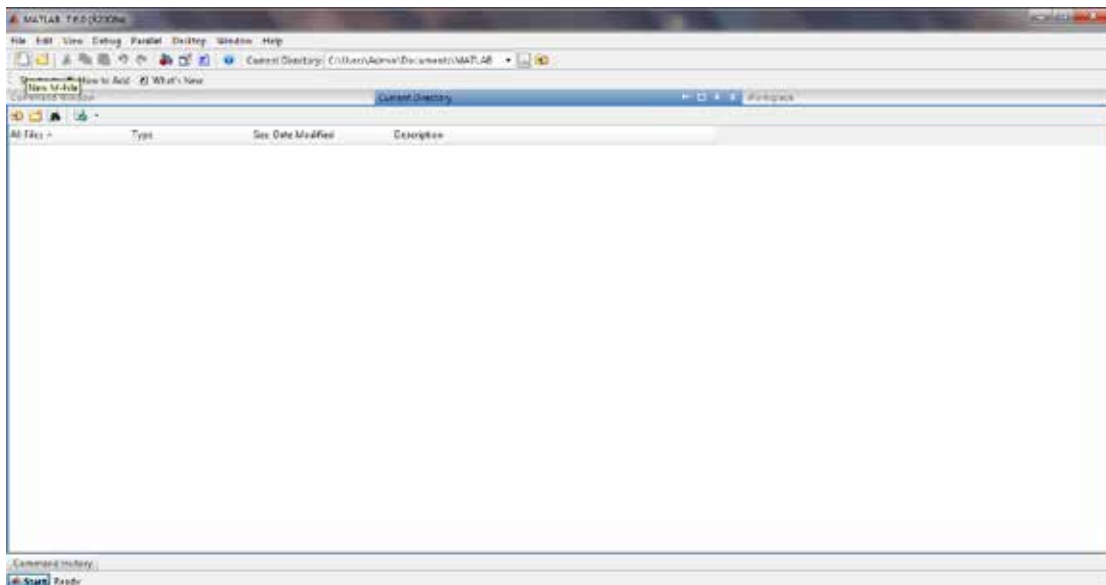
Primeramente, para la base de datos se realizan 5 grabaciones del usuario, las cuales se hicieron con un micrófono en formato .WAV a través de MATLAB a 22 KHz de muestreo y 16 bits de resolución.

Posteriormente se hizo el filtro que elimina el ruido y aumentaba su energía para que las grabaciones sean parecidas a sus parámetros (normalización). Al tener 5 muestras sin ruido y normalizadas se obtiene el promedio de estas para tener una única señal promediada con la señal de voz del usuario.

#### 4.2.1 Creación de la base de datos

MATLAB es un lenguaje de alto nivel y de ambiente interactivo que permite realizar tareas intensas y con una mayor velocidad que los lenguajes de programación comúnmente usados. El lenguaje está construido por un código llamado M-code que puede ser fácilmente ejecutado en la ventana de comando. Con lo cual se pueden crear funciones, etc. Pero la razón principal para la elección de este lenguaje de programación son las herramientas que proporciona para el procesamiento de señales y el conjunto de funciones para el procesamiento digital.

Se inicializa el software, lo cual mostrara la interfaz que se observa en la figura 11, donde se pueden observar varias secciones que indican diferentes opciones, incluyendo la opción de crear un nuevo proyecto, la cual seleccionaremos.



**Figura 11.** Interfaz principal de MATLAB.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Una vez seleccionada la opción de crear un nuevo proyecto. Se despliega una página en blanco donde procederemos a realizar la codificación necesaria para la base de datos como se muestra en la figura 12. Adquirida la señal de voz se procede a eliminar el ruido por medio de un filtro para obtener la señal característica, aunado a esto se debe normalizar la señal, ya que, en MATLAB la señal debe tomar valores desde -1 hasta 1 para poder utilizarla de forma adecuada sin perdidas. Se debe tomar en cuenta que para la grabación de cada audio .wav, se debe modificar el nombre del archivo cada vez que se tenga una nueva grabación.

```

1      %Programa de Grabaciones de la Base de Datos
2 -    Fs = 22050; % Frecuencia de Muestreo
3 -    s = wavrecord(2*Fs, Fs, 1);
4 -    wavwrite(s,Fs,16,'F1.wav') % Guarda .wav señal grabada
5
6      %Programa de Grabaciones de la Base de Datos (Normalización y eliminación
7      %de ruido)
8 -    Fs = 22050; % Frecuencia de Muestreo
9 -    s = wavrecord(2*Fs, Fs, 1);
10 -   wavwrite(s,Fs,16,'F1.wav') % Guarda .wav señal grabada
11 -   lon = length(s); % Longitud del vector
12 -   d = max(abs(s)); % Obtiene el valor más Grande
13 -   s = s/d; % Normaliza la señal
14 -   prom = sum(s.*s)/lon; % Promedio señal entera
15 -   umbral = 0.02; % 2% de la energía promedio
16 -   y = 0;
17 -   for i = 1:400:lon-400 % Ventaneo cada 10ms
18 -       seg = s(i:i+399); % Segmentos
19 -       e = sum(seg.*seg)/400; % Promedio de cada segmento
20 -       % Si el promedio energético es mayor que la señal completa por el valor
21 -       %umbral
22 -       if( e > umbral*prom)
23 -           % Almacena en (y) sino es eliminado como espacio en blanco
24 -           y = [y;seg(1:end)];
25 -       end
26 -   end
27
28 -   % Guarda .audio señal normalizada y sin ruido
29 -   wavwrite(y,Fs,16,'F1b.wav')
30 -   plot(y) % Grafica señal Normalizada
31 -   sound(y,22050) % Reproduce Señal Normalizada

```

**Figura 12.** Programación de la base de datos.

Fuente. Izquierdo, M y Maldonado, A (2019).

Al finalizar la grabación de cada uno de los cinco audios y normalizados, se procede a programar el promedio de cada usuario como se muestra en la figura 13, esto se realiza para obtener una sola señal que caracterizara el comando grabado. Una vez más se debe normalizar la señal.

```

32
33     %Programa que Realiza el Promedio de la Base de Datos de Grabaciones
34 -   clc;
35
36 -   Fs = 22050; %Frecuencia de Muestreo
37 -   rec1 = wavread('F1b.wav'); %Lee la primera grabación
38 -   rec2 = wavread('F2b.wav'); %Lee la segunda grabación
39 -   rec3 = wavread('F3b.wav'); %Lee la tercera grabación
40 -   rec4 = wavread('F4b.wav'); %Lee la cuarta grabación
41 -   rec5 = wavread('F5b.wav'); %Lee la quinta grabación
42
43     %Crea el vector final e grabación
44 -   prom= rec1;
45 -   tam= length(rec5);
46 -   for i= 1:1:tam
47 -       prom(i)=(rec1(i)+rec2(i)+rec3(i)+ rec4(1)+ rec5(i))/5;
48 -   end
49 -   lon = length(prom); %Longitud del vector
50 -   d= max(abs(prom)); %Obtiene el valor más grande
51 -   prom= prom/d; %Normaliza la señal
52 -   %Guarda .wav señal promediada
53 -   wavwrite(prom,Fs,16,'MProm.wav')
54 -   plot(prom) %Grafica señal promediada
55 -   sound(prom,Fs) %Reproduce señal promediada
56

```

**Figura 13.** Programación del promedio de cada usuario.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Obtenidos los promedios de cada usuario, se crea una base de datos con ellos como se muestra en la figura 14, que servirán para comparar con las grabaciones en tiempo real de identificación del software.

```

57     %Base de datos
58 -   [k0,Fs,bits] = wavread('MProm.wav'); %Archivo de promedio Maria
59 -   [k1,Fs,bits] = wavread('AProm.wav'); %Archivo de promedio Alejandro
60 -   [k2,Fs,bits] = wavread('VProm.wav'); %Archivo de promedio Virginia
61

```

**Figura 14.** Base de datos de usuarios.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

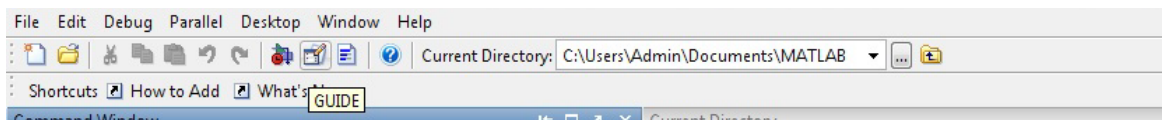
### 4.3 Fase III: Desarrollar un programa con lenguaje de alto nivel que funcione como interfaz de usuario para el reconocimiento de la voz.

El ambiente de desarrollo de interfaz gráfica de usuario (GUIDE, por sus siglas en inglés, graphical user interfase development evironment) es una serie de herramientas que se extiende por completo por el software de MATLAB, diseñada para crear interfaces graficas de usuarios (GUI's, por sus siglas en inglés, graphical user interfaces) de manera fácil, pues no existe la necesidad de que el programador tenga que agregar código para el desarrollo de la misma, debido a que MATLAB automáticamente realiza esta acción, y rápida, ya que presta ayuda en el diseño y presentación de los elementos de las interfaz, reduciendo la labor al grado de seleccionar, tirar, arrastrar y personalizar propiedades. Los elementos que se usar son: botones (push buttons) y ejes (axes).

Una vez que los elementos están en posición se editan las funciones de llamadas (callback) de cada uno de ellos, escribiendo el código de MATLAB que se ejecutara cuando el llamado sea utilizado. Cabe destacar que el orden del programa con el que trabaja la GUI, está determinado por el usuario, pues no existe un flujo establecido como en el caso de un programa común de MATLAB, dándole al usuario la opción de elegir que elementos y en qué momento usarlo, sin ser esto motivo para que la GUI termine su funcionamiento, a menos de que el usuario así lo desee.

#### 4.3.1 Diseño de los componentes (botones y ejes) que formarán la GUI.

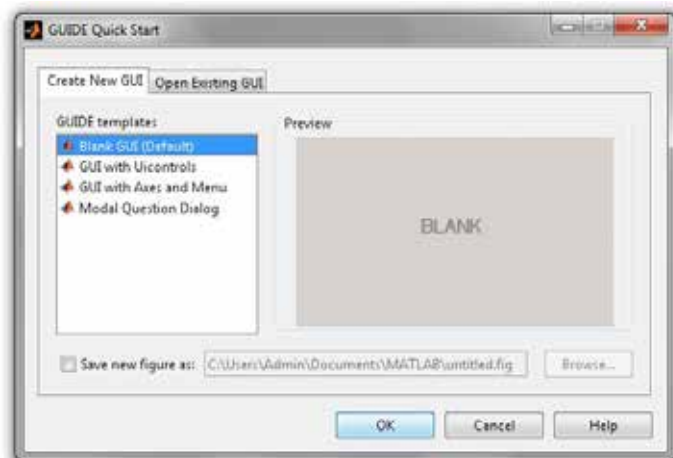
A la herramienta GUIDE la cual se usará para crear la GUI se accede de la manera que se puede ver en la figura 15.



**Figura 15.** Modo de acceso a GUIDE en MATLAB.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

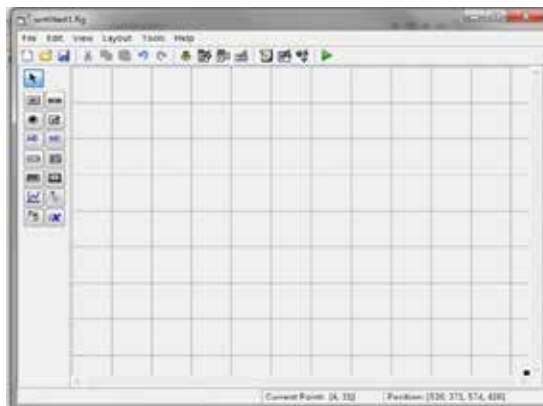
Una vez seleccionada la función GUIDE, aparecerá la ventana de inicio rápido de GUIDE, como se muestra en la figura 16, donde se seleccionará la opción “Blank GUI (Default)” para así iniciar la programación de la GUI.



**Figura 16.** GUIDE Quick Start.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

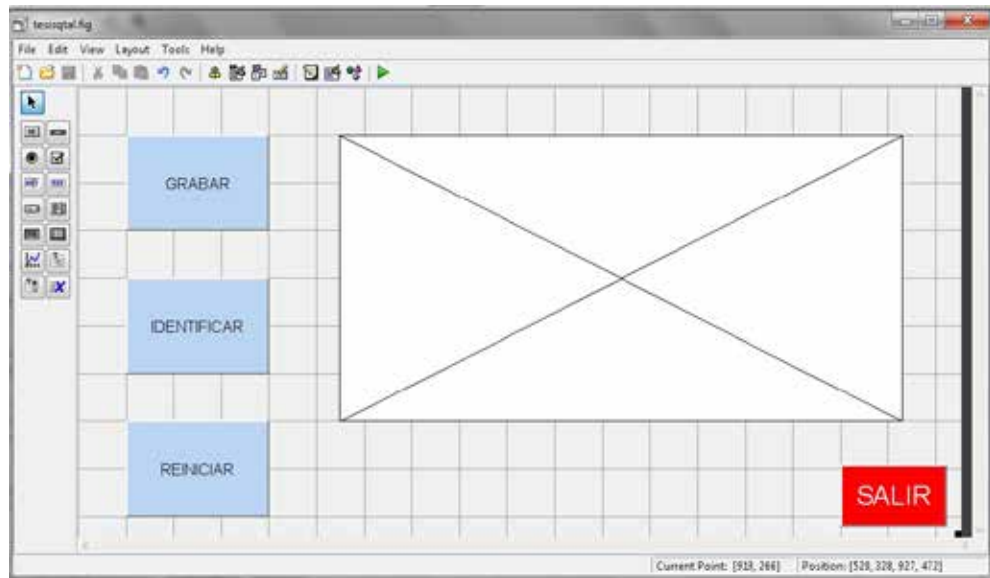
Seguidamente se creará una interfaz en blanco predeterminada como se muestra en la figura 17, donde se diseñará la interfaz conforme a las necesidades del programa a realizar.



**Figura 17.** Espacio de trabajo para crear una interfaz.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Se seleccionan 4 botones que se procederán a editar y colocar el nombre correspondiente (grabar, identificar, reiniciar y salir), seguidamente se colocarán los ejes, donde se mostrarán las gráficas del sistema, todo lo dicho se muestra en la figura 18.



**Figura 18.** Diseño de la interfaz gráfica.  
**Fuente.** Izquierdo, M y Maldonado, A (2019).

#### **4.3.2 Programación de cada uno de los componentes ante la interacción del usuario.**

Se procederá a programar cada uno de los botones de interacción, empezando por la opción de grabar, en la figura 19 se muestra la programación correspondiente a la captación de voz y normalización de la misma.

```

30
31 % --- Boton de grabar
32 function grabar_Callback(hObject, eventdata, handles)
33 - Fs = 22050; % Frecuencia de Muestreo
34 - s = wavrecord(1*Fs, Fs, 1);
35 - wavwrite(s,Fs,16,'F6.wav') % Guarda .wav señal grabada
36
37 %Programa de Grabaciones de la Base de Datos (Normalización y eliminación
38 %de ruido)
39 - Fs = 22050; % Frecuencia de Muestreo
40 - s = wavrecord(1*Fs, Fs, 1);
41 - wavwrite(s,Fs,16,'F6.wav') % Guarda .wav señal grabada
42 - lon = length(s); % Longitud del vector
43 - d = max(abs(s)); % Obtiene el valor más Grande
44 - s = s/d; % Normaliza la señal
45 - prom = sum(s.*s)/lon; % Promedio señal entera
46 - umbral = 0.02; % 2% de la energía promedio
47 - y = 0;
48 - for i = 1:400:lon-400 % Ventaneo cada 10ms
49 -     seg = s(i:i+399); % Segmentos
50 -     e = sum(seg.*seg)/400; % Promedio de cada segmento
51 -     % Si el promedio energético es mayor que la señal completa por el valor
52 -     %umbral
53 -     if (e > umbral*prom)
54 -         % Almacena en (y) sino es eliminado como espacio en blanco
55 -         y = [y;seg(1:end)];
56 -     end
57 - end
58

```

**Figura 19.** Programación para la captación de voz.

Fuente. Izquierdo, M y Maldonado, A (2019).

Seguidamente se muestra en la figura 20 como se guarda la señal limpia y como se crea la gráfica correspondiente.

```

59 % Guarda .wav señal normalizada y sin ruido
60 - wavwrite(y,Fs,16,'Limpia.wav')
61 - plot(y) % Grafica señal Normalizada
62 - sound(y,22050) % Reproduce Señal Normalizada
63 - title ('Grafica de Voz', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize',20,'color', [0 0 0])
64 - xlabel ('Tiempo', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize',15,'color', [0 0 0])
65 - ylabel('Nivel de voz', 'Fontname', 'Dotum', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize',15,'color', [0 0 0])
66 - msgbox('Grabacion terminada');

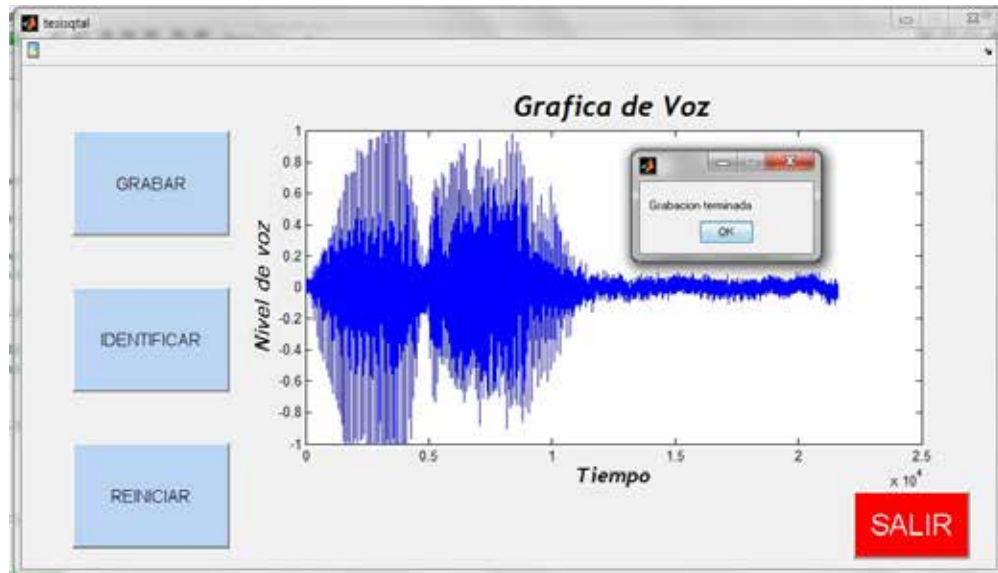
```

**Figura 20.** Programación de la señal limpia y grafica a mostrar.

Fuente. Izquierdo, M y Maldonado, A (2019).

El botón de grabar realiza la función de activar el micrófono y grabar durante 1 segundo el comando a reconocer, en este caso es el nombre del usuario. Para posteriormente limpiar la señal y normalizarla, luego de haber hecho esto imprime una

gráfica de la señal de voz en el recuadro de ejes y aparece una ventana que indica que la grabación ha finalizado, como se observa en la figura 21.



**Figura 21.** Interfaz gráfica luego de haber presionado el botón de grabar.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Al finalizar la grabación se procede a la identificación de la persona que intenta acceder al sistema, primeramente, se programa para obtener los coeficientes LPC de la señal grabada y las 3 señales de la base de datos, como se muestra en la figura 22.

```

69     % --- Boton de identificar
70     function identificar_Callback(hObject, eventdata, handles)
71     [k0,Fs,bits] = wavread('MProm.wav'); %Archivo de promedio Maria
72     [k1,Fs,bits] = wavread('AProm.wav'); %Archivo de promedio Alejandro
73     [k2,Fs,bits] = wavread('VProm.wav'); %Archivo x
74     [kx,Fs,bits] = wavread('Limpia.wav'); %Archio grabado en la GUI
75     load Tesis.mat;
76     num= (Fs/1024)+3; %Obtiene el numero de coeficientes
77     W0= lpc(k0,num); %Obtiene los coeficientes del filtro LPC de Maria
78     W1= lpc(k1,num); %Obtiene los coeficientes del filtro LPC de Alejandro
79     W2= lpc(k2,num); % Obtiene los coeficientes del filtro LPC de Virginia
80     Wx= lpc(kx,num); %Obtiene los coeficientes del filtro LPC de la GUI

```

**Figura 22.** Programación de los LPC.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Seguidamente se obtiene la relación de la señal grabada con cada una de las señales de la base de datos, por medio de la distancia entre cada coeficiente LPC de cada uno de los 4 vectores, como se muestra en la figura 23.

```

81 %Vector que obtendra la distancia entre Maria y la GUI
82 d0= 0;
83 for z=1:25
84     d0(z)= sqrt(((W0(z)-Wx(z))* (W0(z)-Wx(z)))); %Distancia cuadratica media
85     if d0(z)<= 0.15
86         d0(z)=1; %Coloca 1 en el vector si es proximo a Maria
87     else d0(z)= 0; %Coloca 0 en el vector si es lejano a Maria
88     end
89 end
90 %Vector que obtendra la distancia entre Alejandro y la GUI
91 d1= 0;
92 for z=1:25
93     d1(z)= sqrt(((W1(z)-Wx(z))* (W1(z)-Wx(z)))); %Distancia cuadratica media
94     if d1(z)<= 0.15
95         d1(z)=1; %Coloca 1 en el vector si es proximo a Alejandro
96     else d1(z)= 0; %Coloca 0 en el vector si es lejano a Alejandro
97     end
98 end
99 %Vector que obtendra la distancia entre Virginia y la GUI
100 d2= 0;
101 for z=1:25
102     d2(z)= sqrt(((W2(z)-Wx(z))* (W2(z)-Wx(z)))); %Distancia cuadratica media
103     if d2(z)<= 0.15
104         d2(z)=1; %Coloca 1 en el vector si es proximo a x
105     else d2(z)= 0; %Coloca 0 en el vector si es lejano a x
106     end
107 end

```

**Figura 23.** Programación de distancias.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Sigue la etapa donde se da un rango de aceptación en el cual un coeficiente LPC al parecerse a otro se toma en cuenta o se desprecia, al final todos los coeficientes LPC de la relación se suman dando la relación de cuanto se parece la señal grabada con las 3 de la base de datos, programando contadores como se muestra en la figura 24.

```

108 %Contador que almacena el numero de coeficientes que aparecen
109 cont0=0;
110 for i=1:25
111     cont0= d0(i)+cont0; %Suma todos los valores de la cadena
112 end
113 cont1=0;
114 for j= 1:25
115     cont1= d1(j)+ cont1;
116 end
117 cont2=0;
118 for a=1:25
119     cont2= d2(a)+ cont2;
120 end

```

Programación de contadores.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

La etapa final es la toma de decisión la cual consiste en tomar en cuenta 2 condiciones, una que el número coeficientes LPC que se tomaron en cuenta superen un número mínimo requerido para cada usuario y dos que el número de coeficientes LPC no sean iguales, ya que al tener dos relaciones iguales el programa entra en conflicto para decidir que usuario identificar, como se muestra en la figura 25 para uno de los usuarios y en caso de no ser identificado.

```

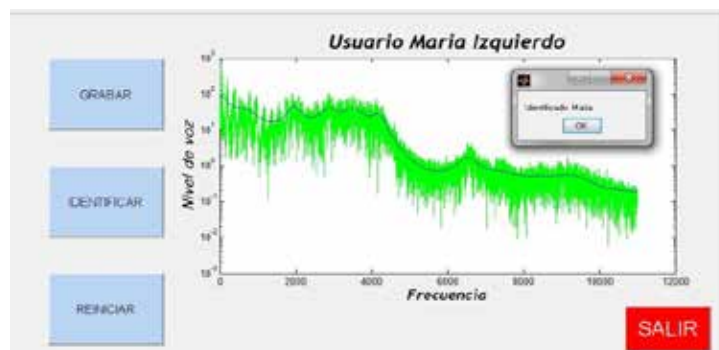
159 - elseif (cont2 > 12)
160 - op= length(k2);
161 - predic= filter([0-W2(2:end)],1,k2); %Prediccion de la señal por Yiltro
162 - error= k2- predic; %Error
163 - Eav= xcorr(k2); %Vector de correlacion
164 - R= Psv(op:op+num); %Obtencion de R(i)
165 - G= sqrt(sum(W2.*R)); %Obtencion de G
166 - envolvente = abs(G./fft(W2,op)); %Obtencion de envolvente H(z)
167 - SW= abs(fft(k2,op)); %Transformada de Fourier de la señal original
168 - semilogy(SW(1:(op/2)), 'g');
169 - hold on;
170 - semilogy(engvolvente(1:(op/2)), 'b');
171 - hold off;
172 - title('Usuario Virginia Flores', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize', 20, 'color', [0 0 0]);
173 - xlabel('Frecuencia', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize', 15, 'color', [0 0 0]);
174 - ylabel('Nivel de voz', 'Fontname', 'Dotum', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'Fontsize', 15, 'color', [0 0 0]);
175 - msgbox('Identificado: Virginia');
176 - Tezis = 3;
177 - save Tezis.mat;
178 - else msgbox('NO IDENTIFICADO');
179 - end

```

**Figura 25.** Programación de la identificación.

Fuente. Izquierdo, M y Maldonado, A (2019).

El botón de identificar realiza todo el proceso explicado anteriormente para así mostrar el nombre de la persona identificada y dar acceso al sistema como se muestra en la figura 26.



**Figura 26.** Interfaz gráfica luego de presionar identificar.

Fuente. Izquierdo, M y Maldonado, A (2019).

En el caso de que la grabación no concuerde con ninguna existente en la base de datos aparecerá un mensaje de no identificado como se muestra en la figura 27.



**Figura 27.** Interfaz gráfica cuando no se es identificado.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

El botón de reiniciar se ocupa de cerrar sesión y volver al inicio de la interfaz, este botón limpia las variables utilizadas y reinicia el programa general, la programación del mismo se muestra en la figura 28.

```

181 % --- Boton de reiniciar.
182 function reiniciar_Callback(hObject, eventdata, handles)
183     close(GCBBF); %Cierra la interfaz grafica
184     run tesisq1tal; %Abre de nuevo
185

```

**Figura 28.** Programación del botón de reiniciar.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

El botón de salir apaga y cierra la interfaz borrando todas las variables utilizadas, la programación de este se muestra en la figura 29.

```

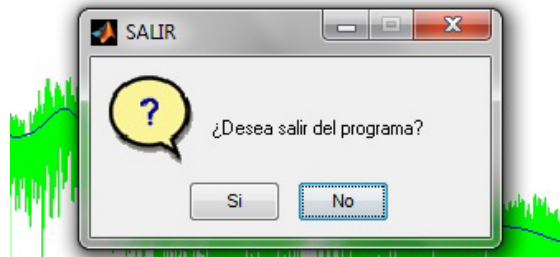
187 % --- Boton de salir.
188 function salir_Callback(hObject, eventdata, handles)
189     ans= questdlg('¿Desea salir del programa?', 'SALIR', 'Si', 'No', 'No');
190     if strcmp(ans, 'No')
191         guidata(hObject, handles);
192         return;
193     end
194     close
195

```

**Figura 29.** Programación del botón de salir.

**Fuente.** Izquierdo, M y Maldonado, A (2019).

Al presionar el botón de salir se muestra una ventana donde se debe seleccionar una acción a realizar, si quedarse dentro del sistema o salir, como se muestra en la figura 30.



**Figura 30.** Interfaz gráfica luego de presionar salir.  
**Fuente.** Izquierdo, M y Maldonado, A (2019).

Por último, se programa la gráfica de voz que tendrá el programa, donde se podrá observar el nivel de voz en función del tiempo, la programación de esta se muestra en la figura 31.

```
193  
196 % --- Grafica del programa  
197 function grafica_CreateFcn(hObject, eventdata, handles)  
198 handles.grafica=1:1:882;  
199 handles.h= plot(handles.grafica,zeros(882,1));  
200 title('Grafica de Voz', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'FontSize',20,'color', [1 1 1]);  
201 xlabel('Tiempo', 'Fontname', 'Trebuchet MS', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'FontSize',15,'color', [1 1 1]);  
202 ylabel('Nivel de Voz', 'Fontname', 'Dotum', 'Fontangle', 'Italic', 'Fontweight', 'Bold', 'FontSize',15,'color', [1 1 1]);
```

**Figura 31.** Programación de la gráfica del programa.  
**Fuente.** Izquierdo, M y Maldonado, A (2019).

#### **4.4 Fase IV: Evaluar el programa de reconocimiento de voz mediante la simulación.**

Se mostrarán los resultados obtenidos al evaluar el programa de reconocimiento de voz, para obtener y comparar los objetivos alcanzados.

Para realizar las siguientes pruebas, se usa la voz de cada uno de los usuarios y dependiendo de los resultados se escogió el mejor lugar para realizar las pruebas de reconocimiento. Dichas pruebas consisten en la comparación de coeficientes LPC de la voz grabada en la base de datos y la voz en tiempo real, para esto existen 3 contadores que ayudarán en la elección.

∅ Cont0 corresponde a María.

Ø Cont1 corresponde a Alejandro.

Ø Cont2 corresponde a Virginia.

Las pruebas realizadas en un lugar cerrado con nivel de ruido de 47 a 55 dBA, se muestran en la tabla 1.

**Tabla 1.** Pruebas con nivel de ruido de 47 a 55 dBA.

Corresponde a:	Cont0	Cont1	Cont2
María	6	20	9
María	5	13	15
Alejandro	15	17	9
Alejandro	13	20	10
Virginia	10	15	13
Virginia	9	16	13

Las pruebas realizadas en un lugar abierto con nivel de ruido de 63 a 70 dBA, se muestran en la tabla 2.

**Tabla 2.** Pruebas con nivel de ruido de 63 a 70 dBA.

Corresponde a:	Cont0	Cont1	Cont2
María	10	8	6
María	11	8	9
Alejandro	7	10	12
Alejandro	6	11	15
Virginia	8	7	13
Virginia	5	6	11

Las siguientes pruebas realizadas se obtuvieron en un lugar con nivel de ruido de 75,6 a 80 dBA, dando como resultado lo mostrado en la tabla 3.

**Tabla 3.** Pruebas con nivel de ruido de 75,6 a 80 dBA.

Corresponde a:	Cont0	Cont1	Cont2
María	17	19	15
María	18	18	13
Alejandro	16	18	10
Alejandro	17	22	17
Virginia	10	10	19
Virginia	17	11	12

El lugar que se escogió para la realización de las pruebas generales fue donde se obtuvieron mejores resultados, mostrado en la tabla 2, donde los niveles de ruido se encuentran entre 63 y 70 dBA.

#### **4.4.1 Pruebas del reconocimiento.**

En las pruebas de reconocimiento, cada usuario grabó su comando de voz con el cual accederá al programa, en estas pruebas se procede a que cada usuario registrado en la base de datos intente acceder mencionando el comando de otro usuario.

Inicialmente los usuarios intentan acceder como “María” y los resultados obtenidos se muestran en la tabla 4.

**Tabla 4.** Resultados intentando acceder como María.

Usuario	Id. María	Id. Alejandro	Id. Virginia	No identificado
María	X			
María	X			
María			X	
María	X			
Alejandro				X
Alejandro				X
Alejandro				X
Alejandro			X	
Virginia				X
Virginia	X			
Virginia				X
Virginia				X

Seguidamente los usuarios intentan acceder como “Alejandro” y los resultados obtenidos se muestran en la tabla 5.

**Tabla 5.** Resultados intentando acceder como Alejandro.

Usuario	Id. María	Id. Alejandro	Id. Virginia	No identificado
María				X
María				X
María				X
María				X
Alejandro		X		
Alejandro		X		
Alejandro		X		
Alejandro				
Virginia				X
Virginia		X		
Virginia				X
Virginia				X

Finalmente, los usuarios intentan acceder como “Virginia” y los resultados obtenidos se pueden observar en la tabla 6.

**Tabla 6.** Resultados intentando acceder como Virginia.

Usuario	Id. María	Id. Alejandro	Id. Virginia	No identificado
María				X
María			X	
María				X
María				X
Alejandro				X
Alejandro				X
Alejandro				X
Alejandro				X
Virginia			X	
Virginia			X	
Virginia			X	
Virginia				X

#### 4.4.2 Pruebas con personas no registradas.

Para estas pruebas se utilizaron 2 sujetos de prueba, un hombre adulto y una chica adolescente, se trató de acceder al sistema como alguno de los usuarios.

Primeramente, se realizaron las pruebas con un hombre adulto intentando acceder como las 3 personas registradas en el sistema, los resultados obtenidos se muestran en la tabla 7.

**Tabla 7.** Resultados para un hombre adulto.

Hombre adulto	Id. María	Id. Alejandro	Id. Virginia	No identificado
María				X
María				X
María				X
María				X
Alejandro				X
Alejandro				X
Alejandro		X		
Alejandro				X
Virginia				X
Virginia				X
Virginia				X
Virginia				X

Seguidamente, se realizaron pruebas con una chica adolescente, intentando acceder como los 3 usuarios registrados en el sistema, los resultados se muestran en la tabla 8.

**Tabla 8.** Resultados para una chica adolescente.

Chica adolescente	Id. María	Id. Alejandro	Id. Virginia	No identificado
María				X
María				X
María				X
María	X			
Alejandro				X
Alejandro				X
Alejandro				X
Alejandro				X
Virginia				X
Virginia			X	
Virginia				X
Virginia				X

## **CAPÍTULO V**

### **CONCLUSIONES Y RECOMENDACIONES**

A continuación, se presentarán las conclusiones más resaltantes del estudio realizado, así como las recomendaciones para futuras investigaciones, con el propósito de avanzar en el diseño de sistemas de seguridad por reconocimiento de voz cada vez más precisos.

#### **5.1 Conclusiones**

- Ø El diseño planteado en este proyecto de investigación ofrece un sistema eficiente para realizar el proceso de reconocimiento de voz.
- Ø El sistema de seguridad por voz es confiable ya que, pasa por diversos métodos siendo el más importante la predicción lineal, obtención de parámetros llamados coeficientes LPC y condiciones implementadas por el programador, disminuyendo la probabilidad de error al reconocer al usuario
- Ø No se podrán agregar usuarios nuevos al instante, tendrán que ser agregados previamente por el proceso de adquisición y entrenamiento por lo cual es un proceso tardado con acreditación del dueño lo cual permite mayor confiabilidad al sistema.
- Ø La interfaz desarrollada permite acceder al sistema de seguridad por reconocimiento de voz, de forma sencilla.
- Ø Se utilizó el diseño en MATLAB ya que este hace que la programación sea sencilla y el acceso de cada usuario con la herramienta GUI es sumamente fácil.

## 5.2 Recomendaciones

- Ø Se propone implementar este sistema de seguridad por reconocimiento de voz en áreas con niveles de ruido no mayores a 70dB, para que el sistema de reconocimiento de voz funcione de modo que exista menor probabilidad de error.
- Ø Realizar una base de datos más extensa de cada usuario con distintos estados de ánimo ya que esto modifica las características de su voz afectando así los coeficientes LPC.
- Ø Utilización de un mejor micrófono debido a que esto permite obtener muestras de mejor calidad y por lo tanto dejar que el software se encargue solo del procesamiento de la señal y no tenga que limpiarla, filtrarla y normalizarla.
- Ø Programar un filtro que elimine el ruido en el que se están tomando las pruebas ya que, esto altera las muestras de la base de datos y la señal adquirida en el momento del reconocimiento de voz.

## REFERENCIAS

### **Bibliográficas**

- Arias, F. (2012). **El proyecto de investigación. Introducción a la metodología científica.** Caracas; Editorial Episteme.
- Balestrini, M (2004). **Como se elabora el proyecto de investigación.** Caracas, Venezuela.
- Hurtado, J. (2010). **Metodología de la investigación.** Caracas, Editorial Quirón.
- Mijares, H. y García, L. (2007) **Manual de Investigación para trabajo especial de grado de la UJAP.** San Diego. Editorial UJAP
- Palella, S y Martins, F (2012). **Metodología de la investigación cuantitativa.** Caracas, Venezuela.
- Sabino, C (1992). **El proceso de investigación.** Editorial Panapo. Caracas, Venezuela.
- Tamayo, M (2003). **El proceso de investigación científica.** Editorial Limusa, Mexico.

### **Electrónicas**

- Caloto, C. (2010). “**Tecnología de reconocimiento por voz y su aplicación en videojuegos**”, Universidad Complutense. Madrid, España.  
<https://eprints.ucm.es/11295/1/SistemasInform%C3%A1ticos.pdf>
- Camargo, J. (2013). “**Sistema de reconocimiento de voz por hardware**”, Universidad Pontificia Bolivariana. Medellín, Colombia.  
[https://repository.upb.edu.co/bitstream/handle/20.500.11912/833/digital\\_19158.pdf?sequence=1](https://repository.upb.edu.co/bitstream/handle/20.500.11912/833/digital_19158.pdf?sequence=1)
- Gómez, J. (2012). “**Apuntes: Medidas de distancia y reconocimiento de vocales**”  
<http://www.eie.fceia.unr.edu.ar/%7Eprodivoz/apuntesindex/pdf>

- Gonzales, J. (2013). “**Reconocimiento de voz con datos perdidos o inciertos**”  
Universidad de Granada. Granada, España.  
<http://digibug.ugr.es/bitstream/handle/10481/29507/21876629.pdf?sequence=1&isAllowed=y>
- Mercado, E. (2018). “**Asociación Internacional de Transporte Aéreo deja Caracas por crisis**”. <https://www.eltiempo.com/mundo/venezuela/iata-abandona-venezuela-por-inestabilidad-politica-158778>
- Yuconza, A. (2017). “**Venezuela. Cifras de la inseguridad**”.  
<https://observatoriodeviolencia.org.ve/venezuela-cifras-de-la-inseguridad/>

