

UNIVERSIDAD MAYOR DE SAN ANDRES
FACULTAD TÉCNICA
CARRERA: ELECTRÓNICA Y TELECOMUNICACIONES



PROYECTO DE GRADO:
SISTEMA DE CONTROL PARA UN EQUIPO DE
TELEVISIÓN BASADO EN EL
RECONOCIMIENTO DE PATRONES DE VOZ
DESTINADO A PERSONAS CUADRAPLÉJICAS

Postulante: Elias Ali Alvarez

Tutor: Lic. Luís Richard Márquez Gonzáles

La Paz- Bolivia

2011

DEDICATORIA

A mis queridos papás Eliodoro y María a quienes amo y respeto, ellos siempre me inculcaron buenos valores, motivándome en todo momento a seguir adelante hasta alcanzar mis metas, apoyándome en todas mis decisiones. Por todo lo anterior este proyecto de grado va dedicado especialmente a mis queridos padres.

A mis hermanos Rita, Thania y Elvis quienes me brindaron su amor y apoyo incondicional.

AGRADECIMIENTOS

A Dios que siempre esta a mi lado para guiarme y cuidarme en los caminos de la vida.

A la Universidad Mayor de San Andrés, por haberme acogido en sus aulas y haberme proporcionado conocimientos.

A la Facultad Técnica y en especial a la carrera de Electrónica y Telecomunicaciones, en la que me forme y adquirí los conocimientos necesarios para desarrollar e implementar el presente proyecto de grado.

Doy las gracias a mis padres Eliodoro Ali y Maria Alvarez, que me apoyaron durante toda mi educación siendo amorosos y comprensivos.

Al Lic. Luis Richard Márquez Gonzáles, quien gracias a sus recomendaciones y correcciones se término con éxito el presente proyecto de grado.

Al Ing. Luis Cabezas Tito, quien me brindo sus conocimientos y amistad sincera.

A mis hermanos Rita, Thania y Elvis, quienes me apoyaron e incentivaron para concluir exitosamente el presente proyecto. Y que en un futuro próximo el presente proyecto sirva de guía y ejemplo para que ellos también concluyan sus estudios con éxito.

RESUMEN

El reconocimiento de voz es el proceso de reconocer automáticamente a una persona en base a la información emitida por su voz. Esta técnica permite hacer uso de la voz como herramienta en la verificación de la identidad de una persona con la finalidad de controlar diversas aplicaciones mediante comandos de voz. Esta tecnología es aprovechada para diversas aplicaciones, en el presente proyecto dicha tecnología se aplica para ayudar a un sector vulnerable de la población que son las personas con discapacidad cuadruplégica.

La cuadruplejía se define como la parálisis bilateral simétrica de ambas extremidades superiores e inferiores.

A lo largo del presente proyecto, se observara el análisis, diseño y la implementación del software de reconocimiento de voz, que a su vez interactúa con hardware externo a la PC mediante una interfaz USB.

La interfaz USB que se implementa en el presente proyecto es de una clase CDC (Communication Device Class), este es el método mas sencillo, porque se emula RS-232 con el USB, con la ventaja de que el PC vera la conexión USB como una conexión COM RS-232 y no requerirá cambios en el software existente.

Para la implementación de la interfaz USB se utiliza un microcontrolador que es el PIC18F4550 este es un PIC de gama alta que posee un modulo USB.

Para el control del televisor se realizo modificaciones al control remoto del mismo, logrando operar el control remoto por comandos de voz.

Con todo lo expuesto anteriormente, el presente proyecto pretende mejorar la calidad de vida de las personas cuadruplégicas, desarrollando para tal fin un sistema de reconocimiento de voz para el control de un equipo de televisión destinado a personas que padecen discapacidad cuadruplégica.

CAPÍTULO 1. GENERALIDADES	11
1.1. INTRODUCCIÓN	11
1.2. ANTECEDENTES	11
1.2.1. ANTECEDENTES DEL TEMA	11
1.2.2. ANTECEDENTES DE TRABAJOS AFINES	15
1.3. PLANTEAMIENTO DEL PROBLEMA	16
1.3.1. DESCRIPCIÓN DEL PROBLEMA	16
1.3.2. PROBLEMA PRINCIPAL	17
1.4. OBJETIVOS	17
1.4.1. OBJETIVO PRINCIPAL	17
1.4.2. OBJETIVOS ESPECÍFICOS	18
1.5. JUSTIFICACIÓN	18
1.5.1. JUSTIFICACIÓN TÉCNICA	18
1.5.2. JUSTIFICACIÓN ECONÓMICA	18
1.5.3. JUSTIFICACIÓN SOCIAL	18
1.6. ALCANCES Y LIMITES	19
1.6.1. ALCANCES	19
1.6.2. LIMITES	19
1.7. APORTES	20
1.7.1. APORTES DEL PROYECTO	20
1.7.2. APORTES ACADÉMICOS	20
1.8. METODOLOGÍAS Y TÉCNICAS A EMPLEARSE	20
1.8.1. MÉTODOS	20
1.8.2. TÉCNICAS	21
CAPÍTULO 2. MARCO TEÓRICO	23
2.1. INTRODUCCIÓN	23
2.2. MARCO CONCEPTUAL	23
2.2.1. RECONOCIMIENTO DE PATRONES	23
2.2.2. ENFOQUES DENTRO DEL RECONOCIMIENTO DE PATRONES	24
2.2.3. MODELO DE SISTEMA DE RECONOCIMIENTO DE PATRONES	25
2.2.4. SISTEMAS BIOMÉTRICOS	26
2.2.5. PARÁMETROS BIOMÉTRICOS	30
2.2.6. MODALIDADES BIOMÉTRICAS	32
2.2.7. GENERACIÓN DE LA VOZ	33
2.2.7.1. TONO	35
2.2.7.2. CANTIDAD	35
2.2.7.3. INTENSIDAD	35
2.2.8. FONEMAS	36
2.2.9. COARTICULACIÓN Y PROSODIA	37
2.2.10. MÉTODOS DE ANÁLISIS DE VOZ	38
2.2.10.1. MUESTREO Y CUANTIFICACIÓN	39
2.2.10.1.1. TEOREMA DE NYQUIST	39
2.2.10.1.2. CUANTIFICACIÓN	40
2.2.10.1.3. CUANTIFICACIÓN UNIFORME	40
2.2.10.1.4. CUANTIFICACIÓN LOGARÍTMICA	40
2.2.10.1.5. CUANTIFICACIÓN NO UNIFORME	41
2.2.10.1.6. CUANTIFICACIÓN VECTORIAL	41
2.2.10.2. ELIMINACIÓN DEL RUIDO	41
2.2.10.3. PRE-ÉNFASIS	42
2.2.10.4. SEGMENTACIÓN	43
2.2.10.4.1. VENTANAS DE ANÁLISIS	43

2.2.11.1.	PREDICCIÓN LINEAL.....	45
2.2.11.2.	CEPSTRUM.....	48
2.2.11.2.1.	CARACTERÍSTICAS.....	51
2.2.12.	MEDIDA DE DISTORSIÓN.....	52
2.2.13.	INTERFAZ USB.....	54
2.2.14.	FUNCIONAMIENTO DE LA INTERFAZ USB.....	54
2.2.15.	CONTROLADOR.....	56
2.2.16.	CONCENTRADORES Ó HUBS.....	56
2.2.17.	PERIFÉRICOS.....	57
2.2.18.	DIAGRAMA DE CAPAS.....	59
2.2.19.	CABLES Y CONECTORES.....	60
2.3.	MARCO APLICATIVO.....	61
2.3.1.	LA MÉDULA ESPINAL NORMAL.....	61
2.3.2.	LA MÉDULA ESPINAL DESPUÉS DE LA LESIÓN.....	64
2.3.3.	LESIÓN PARCIAL Ó TOTAL.....	65
2.3.4.	EL NIVEL DE LA LESIÓN.....	65
2.3.5.	LA PARAPLEJÍA.....	66
2.3.6.	CUADRAPLEJÍA.....	68
CAPÍTULO 3. INGENIERÍA DEL PROYECTO		70
3.1.	INTRODUCCIÓN.....	70
3.2.	REQUERIMIENTOS.....	70
3.2.1.	DESCRIPCIÓN GENERAL.....	70
3.2.1.1.	PERSPECTIVA DEL PRODUCTO.....	70
3.2.1.2.	FUNCIONALIDAD DEL PRODUCTO.....	71
3.2.1.3.	CARACTERÍSTICAS DE LOS USUARIOS.....	71
3.2.1.4.	RESTRICCIONES.....	72
3.2.1.5.	SUPOSICIONES Y DEPENDENCIAS.....	72
3.2.1.6.	EVOLUCIÓN PREVISIBLE DEL SISTEMA.....	72
3.2.2.	REQUISITOS ESPECÍFICOS.....	72
3.2.3.	REQUISITOS COMUNES DE LOS INTERFACES.....	74
3.2.3.1.	INTERFACES DE USUARIO.....	74
3.2.3.2.	INTERFACES DE HARDWARE.....	74
3.2.3.3.	INTERFACES DE SOFTWARE.....	74
3.2.4.	REQUISITOS FUNCIONALES.....	75
3.2.5.	REQUISITOS DEL TELEVISOR.....	75
3.2.6.	REQUISITOS NO FUNCIONALES.....	75
3.2.6.1.	REQUISITOS DE RENDIMIENTO.....	75
3.2.6.2.	FIABILIDAD.....	76
3.2.6.3.	DISPONIBILIDAD.....	76
3.2.6.4.	MANTENIBILIDAD.....	76
3.2.6.5.	PORTABILIDAD.....	76
3.3.	ANÁLISIS Y DISEÑO DEL SOFTWARE DE RECONOCIMIENTO DE VOZ.....	77
3.3.1.	OBTENCIÓN DE INFORMACIÓN.....	78
3.3.2.	ELIMINACIÓN DEL RUIDO.....	79
3.3.3.	FILTRO DE PRE-ÉNFASIS.....	80
3.3.4.	SEGMENTACIÓN.....	80
3.3.5.	COEFICIENTES CEPSTRALES.....	82
3.3.6.	EXTRACCIÓN DE CARACTERÍSTICAS.....	83
3.3.7.	MEDIDA DE DISTANCIA.....	84
3.3.8.	CALCULO DE LA MENOR DISTANCIA.....	85
CAPÍTULO 4. IMPLEMENTACIÓN Y PRUEBAS DEL SISTEMA.....		87
4.1.	INTRODUCCIÓN.....	87
4.2.	ANÁLISIS Y DISEÑO DEL PROTOTIPO.....	87

4.2.1.	MODELO UML	87
4.2.2.	CASOS DE USO	88
4.2.3.	DIAGRAMA DE CLASES	91
4.2.4.	DIAGRAMA DE SECUENCIA	92
4.2.5.	DIAGRAMA DE COLABORACIÓN	93
4.2.6.	DIAGRAMA DE COMPONENTES	94
4.2.7.	DIAGRAMA DE DESPLIEGUE	94
4.2.8.	DISEÑO DE PANTALLAS/INTERFACES	95
4.3.	COMUNICACIÓN USB	98
4.3.1.	USB CDC(<i>Communication Device Class</i>)	98
4.3.2.	PROGRAMA DEL MICROCONTROLADOR PIC18F4550	101
4.4.	PRUEBAS Y RESULTADOS	104
4.4.1.	PRUEBAS DEL SOFTWARE	104
CAPÍTULO 5. ANÁLISIS DE COSTOS		112
5.1.	INTRODUCCIÓN	112
5.2.	FACTIBILIDAD TÉCNICA	112
5.2.1.	COSTOS DE DESARROLLO	112
5.2.1.1.	SOFTWARE	112
5.2.1.2.	HARDWARE	113
5.3.	FACTIBILIDAD OPERATIVA	114
5.4.	COSTO DEL SOFTWARE DE RECONOCIMIENTO DE VOZ	115
5.4.1.	ECUACIONES NOMINALES DE COSTE	115
5.4.2.	ATRIBUTOS DE COSTE	116
5.4.2.1.	ATRIBUTOS DEL PRODUCTO	116
5.4.2.2.	ATRIBUTOS DEL PERSONAL	116
5.4.2.3.	ATRIBUTOS DEL PROYECTO	116
5.4.3.	SIGNIFICADO DE LOS ATRIBUTOS	117
5.4.4.	CÁLCULO DEL ESFUERZO, TIEMPO Y NÚMERO DE PERSONAS	122
5.4.5.	ESTIMACIÓN DEL COSTO DEL SOFTWARE	123
5.5.	COSTO TOTAL DEL PROYECTO(CTS)	124
CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES		126
6.1.	CONCLUSIONES	126
6.2.	RECOMENDACIONES	127
ANEXO A	ÁRBOL DE PROBLEMAS	129
ANEXO B	ÁRBOL DE OBJETIVOS	130
ANEXO C	FORMULARIO DE EVALUACIÓN DE PRUEBAS	131
ANEXO D	CARACTERÍSTICAS DEL PIC 18F4550	132
REFERENCIAS	BIBLIOGRÁFICAS	134

ÍNDICE DE FIGURAS**PÁGINA**

Figura 2.1	Etapas en un sistema de reconocimiento de patrones.....	25
Figura 2.2	Sistema Biométrico Genérico.....	28
Figura 2.3.	Proceso de captura y verificación de usuario.....	29
Figura 2.4	Definición de la tasa de error igual.....	32
Figura 2.5	Mercado de Biométricos por tecnología.....	32
Figura 2.6	Aparato fonatorio.....	33
Figura 2.7	Modelo de producción de voz basado en LPC.....	46
Figura 2.8	Modelo de la técnica Homomórfica.....	48
Figura 2.9	Modelo de Coeficientes Cepstrales	49
Figura 2.10	Estructura de capas del bus USB.....	55
Figura 2.11	Esquema de un concentrador.....	56
Figura 2.12	Posible esquema de conexiones del bus USB.....	57
Figura 2.13	Dispositivos USB conectados a un PC.....	58
Figura 2.14	Diagrama de Capas.....	58
Figura 2.15	Esquema del cable para USB.....	60
Figura 2.16	Medula espinal.....	61
Figura 2.17	División de la medula espinal.....	62
Figura 2.18	Las vértebras.....	63
Figura 2.19	Persona parapléjica.....	67
Figura 3.1	Esquema general del proyecto.....	70
Figura 3.2	Esquema del algoritmo de reconocimiento de voz.....	77
Figura 3.3	Diagrama de flujo de la función eliminación del ruido.....	79
Figura 3.4	Diagrama en bloques de un filtro.....	80
Figura 3.5	Diagrama de flujo de la función segmentación.....	81
Figura 3.6	Diagrama de flujo de la extracción de coeficientes ceptrales.....	82
Figura 3.7	Diagrama de flujo de la extracción de características.....	83
Figura 3.8	Diagrama de flujo de la medida de distancia.....	84
Figura 3.9	Diagrama de flujo del calculo de la menor distancia.....	85
Figura 4.1	Diagrama de casos de uso general.....	87
Figura 4.2	Diagrama de clases.....	90
Figura 4.3	Grabación de comandos.....	91
Figura 4.4	Reconocimiento de comando de voz.....	91
Figura 4.5	grabación de comandos.....	92
Figura 4.6	Reconocimiento de comando de voz.....	92
Figura 4.7	Diagrama de componentes.....	93
Figura 4.8	Diagrama de despliegue.....	94
Figura 4.9	Pantalla principal grabación del comando.....	95
Figura 4.10	Pantalla principal mostrando el espectro de frecuencias.....	97
Figura 4.11	Pantalla de base de datos.....	97
Figura 4.12	Circuito simulado en ISIS.....	100
Figura 4.13	Circuito de control simulado en ISIS.....	101
Figura 4.14	Circuito real armado en un protoboard.....	103
Figura 5.1	prueba de caja negra.....	110

ÍNDICE DE TABLAS	PÁGINA
Tabla 2.1 Clasificación de los fonemas en castellano.....	36
Tabla 2.2 Clasificación de los fonemas vocálicos.....	36
Tabla 3.1 Características de los usuarios.....	71
Tabla 3.2 Requisito Especifico 1.....	73
Tabla 3.3 requisito Especifico 2.....	73
Tabla 3.4 requisito Especifico 3.....	73
Tabla 3.5 Valores de parámetros.....	81
Tabla 4.1 Caso de uso #1.....	88
Tabla 4.2 Caso de uso #2.....	89
Tabla 5.1 Pruebas de reconocimiento de voz.....	104
Tabla 5.2 Resultados de los cuestionarios.....	109
Tabla 6.1 Software para el desarrollo.....	113
Tabla 6.2 Hardware para el desarrollo.....	113
Tabla 6.3 Coeficientes COCOMO Intermedio.....	116
Tabla 6.4 Valores de los atributos.....	117
Tabla 6.5 Salario de un programador por mes.....	123

ÍNDICE DE ECUACIONES	PÁGINA
Ec. 2.1 tasa de falsa aceptación.....	31
Ec. 2.2 tasa de falso rechazo.....	31
Ec. 2.3 teorema de nyquist.....	38
Ec. 2.4 teorema de muestreo.....	39
Ec. 2.5 energía promedio de cada segmento.....	41
Ec. 2.6 energía promedio de la señal entera.....	41
Ec. 2.7 filtrado paso alto.....	42
Ec. 2.8 pre-énfasis.....	42
Ec. 2.9 segmentación.....	43
Ec. 2.10 Ventana rectangular.....	43
Ec. 2.11 Ventana de Hanning.....	43
Ec. 2.12 Ventana de Hamming.....	43
Ec. 2.13 Ventana de Bartlett.....	43
Ec. 2.14 Ventana de Blackman.....	43
Ec. 2.15 aproximación de predicción lineal.....	45
Ec. 2.16 LPC.....	45
Ec. 2.17 Transformada Z de LPC.....	46
Ec. 2.18 función de transferencia de LPC.....	46
Ec. 2.19 error LPC.....	47
Ec. 2.20 Estimación de LPC.....	47
Ec. 2.21 ecuación Matricial LPC.....	47
Ec. 2.22 ecuación simplificada de LPC.....	47
Ec. 2.23 señal de voz en el dominio de la frecuencia.....	49
Ec. 2.24 cepstrum.....	49
Ec. 2.25 cepstrum simplificado.....	49

Ec. 2.26 propiedad de logaritmo.....	50
Ec. 2.27 coeficientes cepstrales	50
Ec. 2.28 representación ortogonal polinomial.....	51
Ec. 2.29 coeficientes representación ortogonal polinomial.....	51
Ec. 2.30 Distancia L_1.....	52
Ec. 2.31 Distancia euclidea L_2	52
Ec. 2.32 distancia Euclidea ponderada	52
Ec. 2.33 distancia de Mahalanobis	52
Ec. 5.1 ecuaciones de COCOMO.....	115



CAPÍTULO I

GENERALIDADES

CAPÍTULO 1. GENERALIDADES

1.1. INTRODUCCIÓN

A través del tiempo, el hombre en su afán por simplificar sus tareas cotidianas, ha desarrollado métodos y herramientas que le han permitido, mejorar su calidad de vida. Como resultado de esa búsqueda constante e incansable, se ha realizado grandes avances científicos y tecnológicos en diversos campos del saber, dentro de los cuales podemos destacar el procesamiento digital de señales específicamente el reconocimiento de patrones de voz.

Estas tecnologías, de manera general, se pueden englobar en un solo objetivo, crear canales de comunicación adecuados a las capacidades de las personas. Algunas de estas personas se ven con la problemática de no contar con ningún tipo de movimiento físico para poder comunicarse, como es el caso de las personas cuadrapléjicas que están aisladas de su entorno a pesar de que su habla goza de total capacidad, y es torno a estas características que se ha buscado dentro de un área muy específica, encontrar una solución a esta problemática, lo cual nos conduce al reconocimiento de voz.

El reconocimiento de voz es el proceso de reconocer automáticamente a una persona en base a la información emitida por su voz. Esta técnica permite hacer uso de la voz como herramienta en la verificación de la identidad de una persona con la finalidad de controlar diversas aplicaciones mediante comandos de voz.

1.2. ANTECEDENTES

1.2.1. ANTECEDENTES DEL TEMA

La tecnología a lo largo de los años ha evolucionado de forma significativa en lograr que el esquema de comunicación de los seres humanos haya conseguido una representación efectiva a través de la conjunción de una

aplicación informática y su desarrollo en una interfaz sencilla de manejar, para que el usuario final pueda llevar a cabo las tareas cotidianas desde su casa y simplemente utilizando la voz como plataforma para la consecución de su objetivo final.

A continuación aparecen las fechas más significativas, así como las distintas investigaciones desarrolladas en el campo del reconocimiento de voz.¹

- **1870's Alexander Graham Bell:** quiso construir un sistema/dispositivo que hiciera el habla visible a las personas con problemas auditivos. El resultado fue el teléfono.
- **1880's Tihmir Nemes:** solicita un patente para desarrollar un sistema de transcripción automática que identificara secuencias de sonidos y los imprimiera (texto). Este proyecto fue rechazado como “proyecto no realista”.
- **30 años después AT&T Bell Laboratorios:** construye la primera maquina capaz de reconocer voz (basada en Templates) de los 10 dígitos del Ingles. Requería extenso reajuste a la voz de una persona, pero una vez logrado tenía un 99% de aciertos. Por lo tanto, surge la esperanza de que el reconocimiento de voz resulte simple y directo.
- **A mediados de los 60's :** la mayoría de los investigadores reconoce que era un proceso mucho mas intrincado y sutil de lo que habían anticipado.

Empiezan a reducir los alcances y se enfocan a sistemas más específicos:

- Dependientes del locutor.
- Flujo discreto de habla (con espacios / pausa entre palabras).
- Vocabulario pequeño(menor o igual a 50 palabras)

¹ Fuente: <http://www.dei.uc.edu.py/tai2000/reconocedor/Historia.htm>

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 de voz continuo pero no se ven resultados hasta la década de los 70.

- **1970-1980:** nace el primer producto de reconocimiento de voz, el **VIP100** de **Threshold Technology Inc.** Utilizaba un vocabulario pequeño, dependiente del locutor, y reconocía palabras discretas. Asimismo, gana el **U.S. National Award en 1972.** Nace el interés de **ARPA**, organismo que pertenece al Departamento de Defensa de los Estados Unidos de América, por los que nos precipitamos a la época de la inteligencia artificial. El proyecto financiado por esta institución buscó el reconocimiento del habla continua, de la ampliación del vocabulario. Impulsa a los investigadores para que se centren en el entendimiento del habla. Los sistemas empiezan a incorporar los siguientes módulos:

- " Análisis léxico (conocimiento léxico)
- " Análisis sintáctico (estructura de palabras)
- " Análisis semántico
- " Análisis pragmático

Este proyecto termina en 1976 con el resultado esperado y las empresas contratadas CMU, SRI y MIT, por medio de sus investigadores crean los siguientes sistemas para ARPA, organismo que pertenece al Ministerio de Defensa de los Estados Unidos.

A) CMU:

- " Harpy
- " Dragon

B) HWIM:

- " Hearsay II -> Votan/Dragon Systems->PC

En la década de los 80 será IBM la que remonta el vuelo. No se olvide que es la empresa que inventa el ordenador personal en agosto de 1981.

“IBM desarrolla N-grams, lo cual forma la base de la mayoría de los sistemas actuales comerciales.

En estos mismos años, surgen los sistemas de vocabulario amplio, que ahora son la norma. (>1000 palabras). Adicionalmente bajan los precios de estos sistemas.

Empresas importantes actualmente que desarrollan aplicaciones de reconocimiento de voz

- Philips
- Lernout & Hauspie
- Sensory Circuits
- Dragon Systems
- Speechworks
- Vocalis
- Dialogic
- Novell
- Microsoft
- NEC
- Siemens
- Intel (apoyo / soporte técnico)

Por otro lado, es necesario recordar que los siguientes productos de reconocimiento de voz son utilizados por la empresa Microsoft:

- “Dragon Dictate and Naturally speaking” de Dragon Systems

Los siguientes productos son compatibles con los sistemas Dragon para asegurar la alta fidelidad del sonido.

- Freespeech by Philips
- Kolvox Lawtalk 2.0 para Windows by Kolvox
- Learout and Hauspie V R technology
- Metroplex Voice computing: desarrollo hands free programs, los cuales utilizan Dragon Systems Speech recognition para dictar matemáticas
- “Voice Pilot” para Windows by Voice Pilot, Inc.
- Via Voice para Windows de IBM

Grabadores digitales

- Dragon Naturally Mobile
- Olympus D1000 digital recorder
- Norcom Dictation Systems
- Sony MZ-R55
- Voice it Products (este producto esta asociado para su uso con los sistemas Dragon)

*Nota: el primer software que se diseñó con el sistema de reconocimiento de voz para PC, con la función de dictado, la desarrolló la empresa Dragon Systems (Dragon Dictate for Winows 1.0), en el año 1994. En 1996, IBM, decidió sacar su propio software de reconocimiento llamado MedSpeak/Radiology. Con motivo de la competencia surgida, Dragon lanza al mercado en junio de 1997 Naturally Speaking e IBM responde con ViaVoice .

1.2.2. ANTECEDENTES DE TRABAJOS AFINES

Realizando una investigación exploratoria en las universidades de la ciudad de La Paz se llego a encontrar los siguientes trabajos de grado que tienen afinidad al presente trabajo.

El primer trabajo es de la EMI (Escuela Militar de Ingeniería), cuya autora es Gabriela Agudo Vera, el trabajo que realizo titula “Sistema basado en redes

neuronales artificiales para la identificación de huellas dactilares en procesos de investigación en escenarios de incertidumbre. Caso de la FELCC”, el mencionado trabajo tiene por objetivo desarrollar e implementar un sistema basado en redes neuronales artificiales para la identificación de huellas dactilares en procesos de investigación en escenarios de incertidumbre. Caso FELCC, que fue presentado el año 2008.

El siguiente trabajo también proviene de la EMI, cuyo autor es Alexander Humberto García García, el trabajo que realizó titula “Diseño de una interfaz hombre maquina aplicando el acelerómetro LIS3LV02DQ para personas con trastornos motores osteoarticulares en las manos”, el mencionado trabajo tiene por objetivo diseñar una interfaz hombre maquina en un sistema basado en microcontroladores aplicando el acelerómetro LIS3LV02DQ para personas con trastornos motores osteoarticulares en las manos, y fue presentado el año 2008.

El ultimo trabajo proviene de la USB (Universidad Salesiana de Bolivia), cuya autora es Dímelas Paola Salazar Apaza, el trabajo que realizó titula “Diseño de una arquitectura de control domótico que brinde confort y seguridad a minusválidos físicos.”, el mencionado trabajo tiene por objetivo diseñar un prototipo de una arquitectura de control domótico que brinde seguridad y confort a minusválidos físicos, y fue presentado el año 2007.

La diferencia del presente proyecto de grado con los anteriores mencionados es que se diseña e implementa un sistema de control para un equipo de televisión, mediante el reconocimiento de patrones de voz, para discapacitados cuadrapléjicos.

1.3. PLANTEAMIENTO DEL PROBLEMA

1.3.1. DESCRIPCIÓN DEL PROBLEMA

Las personas que padecen discapacidad cuadrapléjica o tetrapléjica tienen una parálisis parcial o total de sus miembros superiores e inferiores, lo cual

hace que las mismas no sean autónomas ni independientes en su desenvolvimiento diario, por ejemplo no les es posible hacer uso de los equipos electrónicos que posean en su hogar, tal es el caso del televisor que es uno de los equipos de mayor uso dentro del hogar.

A raíz de esto los discapacitados son personas que precisan de las nuevas tecnologías para poder llevar a cabo su vida, y sobre todo, de la manera más cómoda posible. También se tienen los problemas económicos ya que las mismas no pueden trabajar y generar sus propios recursos, lo cual incide en el presupuesto familiar.

1.3.2. PROBLEMA PRINCIPAL

Luego de un análisis y estudio previo se ha diseñado el árbol de problemas lo que ha permitido identificar el problema principal. (Ver Anexo A)

Las personas con discapacidad cuadrapléjica necesitan operar en forma independiente su equipo de televisión y no lo pueden hacer debido a que sufren parálisis en brazos, piernas o ambos miembros.

1.4. OBJETIVOS

Luego de un análisis y estudio previo se ha diseñado el árbol de objetivos lo que ha permitido identificar los objetivos que persigue el presente proyecto. (Ver Anexo B)

1.4.1. OBJETIVO PRINCIPAL

Implementar previo diseño un sistema de control para un equipo de televisión destinado a personas con discapacidad cuadrapléjica, aplicando reconocimiento de patrones de voz.

1.4.2. OBJETIVOS ESPECÍFICOS

- Investigar técnicas de modelado y diversas técnicas de reconocimiento de voz.
- Diseñar los algoritmos de reconocimiento de voz.
- Diseñar el sistema electrónico que cumpla con el objetivo del presente trabajo.
- Implementar el software de reconocimiento de voz en un lenguaje de programación.
- Implementar el sistema electrónico.
- Realizar la Interacción entre el software y hardware del sistema.
- Diseñar una interfaz amigable para el usuario final.
- Realizar las pruebas y correcciones pertinentes.

1.5. JUSTIFICACIÓN

1.5.1. JUSTIFICACIÓN TÉCNICA

Se justifica técnicamente debido a que en el presente proyecto se innovará la forma de control de un equipo de televisión ya que el control se realizara mediante comandos de voz y así contribuir con el avance de la tecnología.

1.5.2. JUSTIFICACIÓN ECONÓMICA

La justificación económica la analizaremos desde dos puntos de vista, la inversión a realizar para llevar a cabo el proyecto será mínima y los beneficios que se obtendrán para los discapacitados tetraplégicos serán mayores.

1.5.3. JUSTIFICACIÓN SOCIAL

Se justifica socialmente debido a que con el presente proyecto se pretende ayudar a mejorar la calidad de vida de un sector vulnerable de la población que son los discapacitados cuadruplégicos, mediante el presente proyecto

se pretende que las personas con discapacidad cuádruplégica sean más autónomas e independientes. Con todo lo expuesto anteriormente se pretende que la vida de estas personas mejore en varios aspectos.

1.6. ALCANCES Y LIMITES

1.6.1. ALCANCES

A Continuación se detallan los alcances del proyecto:

- Control de un equipo de televisión mediante comandos de voz, que sea usado preferentemente por las personas cuádruplégicas, ya que ellas no pueden acercarse físicamente al equipo de televisión.
- Proporcionar una interfaz amigable para el usuario final, para que el sistema pueda ser operado por cualquier persona que posea el habla en buenas condiciones pero con preferencia por la personas con discapacidad cuádruplégica ya que ellas tienen más dificultades para controlar su equipo de televisión.

1.6.2. LIMITES

A continuación se detallan los límites del proyecto:

- El sistema de control para un equipo de televisión basado en el reconocimiento de patrones voz destinado a discapacitados cuádruplégicos, se limitará a controlar los siguientes eventos: encendido, apagado del televisor, siguiente y anterior canal, aumentar y disminuir volumen.
- El sistema de control para un equipo de televisión destinado a discapacitados cuádruplégicos se limitará al reconocimiento de comandos de voz en forma independiente de quien lo diga, es decir no hará el reconocimiento de locutor.
- Se hará un prototipo para un equipo de televisión en específico.

1.7. APORTES

1.7.1. APORTES DEL PROYECTO

El principal aporte del proyecto es el de proporcionar un software de reconocimiento de patrones de voz que interactúe con hardware de la computadora para el control de un equipo de televisión para así proporcionar una mejor calidad de vida a los discapacitados cuadrapléjicos, debido a que le proporcionara el control de su equipo de televisión mediante comandos de voz, con lo cual la persona discapacitada ya no tendrá la dificultad de operar manualmente su televisor.

En resumen los aportes del proyecto son el de proporcionar mayor autonomía e independencia a las personas con discapacidad cuadrapléjica y que estas personas puedan tener una mejor calidad de vida.

1.7.2. APORTES ACADÉMICOS

El aporte académico es el de realizar un prototipo software que conjuncione el procesamiento digital de señales, reconocimiento de patrones de voz y que interactúe con elementos hardware, logrando así una interfaz humano-maquina, que pueda ser aprovechada preferentemente por las personas con discapacidad cuadrapléjica.

Otro de los aportes del proyecto de grado también es el de mejorar las técnicas de reconocimiento de voz. Sentando precedentes para futuras investigaciones o trabajos de grado que tengan como base el presente trabajo.

1.8. METODOLOGÍAS Y TÉCNICAS A EMPLEARSE

1.8.1. MÉTODOS

Se aplicaron los siguientes métodos:

- Método científico, se utilizó para mejorar y precisar teorías previas en función de nuevos conocimientos.
- Método descriptivo, para describir las herramientas y componentes que se utilizaron.
- Métodos de análisis de voz, estos métodos nos proporcionarán un modelo matemático de la señal de voz [Ver marco teórico]
 - Transformada de Fourier
 - Filtros Digitales
 - Predicción Lineal
 - Modelo LPC(Coeficientes de predicción lineal)
 - Modelo Cepstrum
- Modelos de reconocimiento de voz, estos modelos proporcionarán algoritmos para reconocer la voz.[Ver marco teórico]
 - Cuantización Vectorial
- Modelo UML, para especificaciones, visualización, construcción y documentación de los artefactos del software.
- Modelo Cocomo intermedio, para proporcionar estimaciones del esfuerzo y costos requeridos para llevar a cabo el proyecto.
- Métodos de prueba: caja negra, para la prueba del software.

1.8.2. TÉCNICAS

Las técnicas que se aplicaron para el desarrollo del presente proyecto son las siguientes:

- a) Técnicas de recopilación de datos, para recabar información referida al proyecto.
- b) Técnicas de muestreo, estas técnicas permitieron discernir la información obtenida.



CAPÍTULO II

MARCO TEÓRICO

CAPÍTULO 2. MARCO TEÓRICO

2.1. INTRODUCCIÓN

El proceso de reconocimiento de patrones voz para el control de un equipo de televisión, requiere como cualquier otra investigación científica, los respectivos fundamentos teóricos para su realización, ya que mediante la recopilación de información en el presente capítulo, será posible tener herramientas de análisis para la resolución del problema que se está planteando.

El presente trabajo se basa fundamentalmente en el reconocimiento de patrones de voz, por lo que en el presente capítulo hará más énfasis en los modelos y métodos de reconocimiento de voz.

La tecnología del reconocimiento de patrones de voz es un amplio conjunto de conocimientos y procedimientos sobre la información representada en la señal de voz y que pertenecen a un marco científico-técnico multidisciplinario, donde se dan cita diferentes ramas del saber como son: fisiología, acústica, lingüística, procesamiento digital de señales, técnicas de inteligencia artificial, etc.

2.2. MARCO CONCEPTUAL

2.2.1. RECONOCIMIENTO DE PATRONES²

- Es la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos y/o abstractos, con el propósito de extraer información que permita establecer propiedades de ó entre conjuntos de dichos objetos.
- Es un área dentro de la inteligencia artificial, trabaja en la solución de tareas tales como el reconocimiento de imágenes, reconocimiento del habla. Sin embargo un objetivo mas ambicioso ha sido durante todo el tiempo

² Stanford University

implementar la percepción artificial, es decir, imitar las funciones de los sistemas sensoriales biológicos en su forma mas completa.

- Reconocimiento de patrones es un área de la tecnología conocido como aprendizaje de maquinas (Machine Learning) o aprendizaje automático. El único propósito de este método es el clasificar un grupo de patrones conocido como conjunto de pruebas en dos o más clases de categorías. Esto es logrado al calcular las categorías del conjunto en prueba comparándolo con un conjunto de entrenamiento. Un clasificador dado mide la distancia entre varios puntos dados, para saber cuales puntos son mas cercanos a la meta en un modelo parametrizado.

2.2.2. ENFOQUES DENTRO DEL RECONOCIMIENTO DE PATRONES

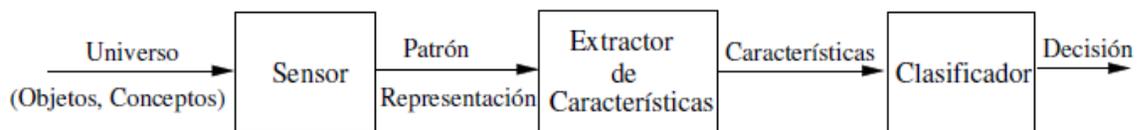
- Geométrico (Clustering): Los patrones deben ser graficables, en éste enfoque se emplea el cálculo de distancias, geometría de formas, vectores numéricos, puntos de atracción, etc.
- Estadístico: Se basa en la teoría de la probabilidad y la estadística, utiliza análisis de varianzas, covarianzas, dispersión, distribución, etc.
- Sintáctico Estructural: estudia la estructura de los objetos, es decir, usa teoría de lenguajes formales, gramáticas, teoría de autómatas, etc.
- Neuro Reticular: Se utilizan redes neuronales que se 'entrenan' para dar una cierta respuesta ante determinados valores.
- Lógico Combinatorio: se basa en la idea de que la modelación del problema debe ser lo más cercana posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas. Se utiliza para conjuntos difusos y utiliza lógica simbólica, circuitos combinacionales y secuenciales, etc.

2.2.3. MODELO DE SISTEMA DE RECONOCIMIENTO DE PATRONES

Los procesos perceptuales del ser humano pueden ser modelados como un sistema de tres estados:

- adquisición de datos sensoriales
- extracción de características
- toma de decisiones

Figura 2.1 Etapas en un sistema de reconocimiento de patrones



Fuente [J. Kittler]

Por lo tanto es conveniente dividir el problema del reconocimiento automático de una manera similar

- Sensor** Su propósito es proporcionar una representación factible de los elementos del universo a ser clasificados. Es un sub-sistema crucial ya que determina los límites en el rendimiento de todo el sistema.

Idealmente uno debería entender completamente las propiedades físicas que distinguen a los elementos en las diferentes clases y usar ese conocimiento para diseñar el sensor, de manera que esas propiedades pudieran ser medidas directamente. En la práctica frecuentemente esto es imposible porque:

- no se dispone de ese conocimiento

- muchas propiedades útiles no se pueden medir directamente (medición no intrusiva)
 - no es económicamente viable
- ii) Extracción de Características** Esta etapa se encarga, a partir del patrón de representación, de extraer la información discriminativa eliminando la información redundante e irrelevante. Su principal propósito es reducir la dimensionalidad del problema de reconocimiento de patrones.
- iii) Clasificador** Es la etapa de toma de decisiones en el sistema. Su rol es asignar a la categoría apropiada los patrones de clase desconocida a priori.

2.2.4. SISTEMAS BIOMÉTRICOS

El término biometría viene del griego “bio” que significa vida y “metría” que significa medida o medición, de acuerdo al diccionario de la real academia de la lengua española biometría es el estudio mensurativo o estadístico de los fenómenos o procesos biológicos, sin embargo más recientemente y para el tema que nos concierne el significado de biometría es el conjunto de métodos automatizados que analizan determinadas características humanas para identificar o autenticar personas.

La biometría aprovecha que hay ciertas características biológicas o conductuales singulares e inalterables, por lo que pueden ser analizados y medidos para crear una huella biométrica. Estas características son difíciles de perder, transferir u olvidar y son perdurables en el tiempo.

La biometría se soporta en siete pilares o conceptos básicos que son:

- Universalidad: que tan común es encontrar este biométrico en los individuos.

- Singularidad: que tan único o diferenciable es la huella biométrica entre uno y otro individuo.
- Permanencia: que tanto perdura la huella biométrica en el tiempo de manera inalterable.
- Recolectable: Que tan fácil es la adquisición, medición y almacenamiento de la huella biométrica.
- Calidad: que tan preciso, veloz y robusto es el sistema en el manejo de la huella biométrica.
- Aceptabilidad: Que tanta aprobación tiene la tecnología entre el público.
- Fiabilidad: Que tan fácil es engañar al sistema de autenticación.

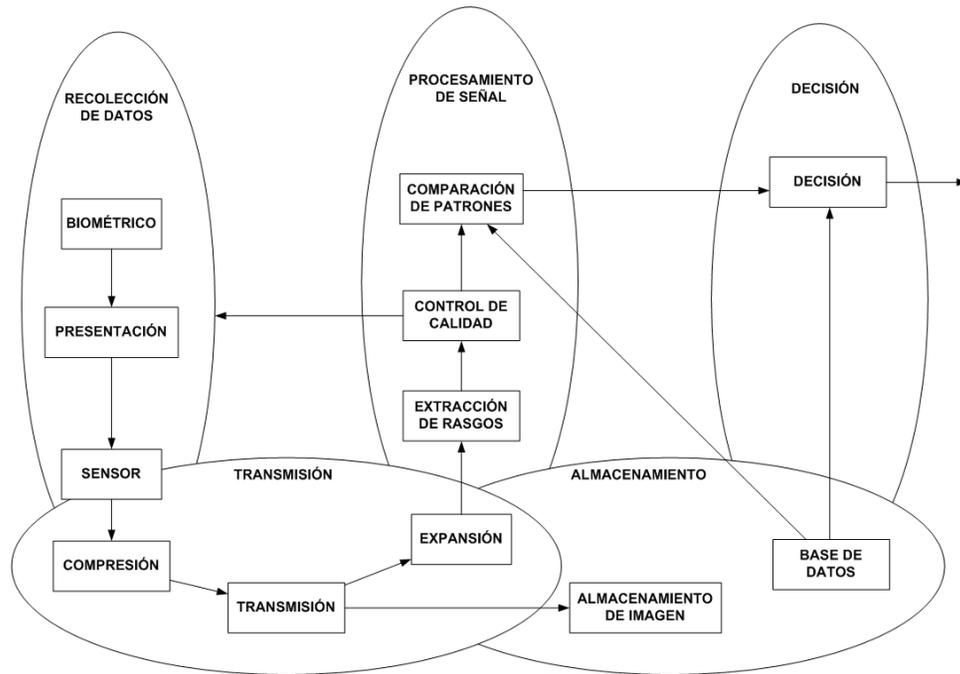
En la biometría se distinguen dos grupos de registros biométricos los fisiológicos o morfológicos y los conductuales.

Los biométricos morfológicos o fisiológicos son aquellos que se soportan sobre características físicas inalterables y presentes en la mayoría de los seres humanos tales como: huella dactilar, la voz, geometría de la mano, características del iris, patrones vasculares de la retina, mano, etc.

Los biométricos conductuales son aquellos que se soportan sobre características de la conducta del ser humano tales como: pulsaciones del teclado, discurso, dinámica de la firma, etc.

En general un sistema biométrico se puede esquematizar de la siguiente manera:

Figura 2.2 Sistema Biométrico Genérico



Fuente [Fundamentals of Biometric Authentication Technologies, James I. Wayman]

En la biometría hay tres términos de uso muy frecuente que son reconocimiento, verificación e identificación, cada uno de estos términos que a simple vista parecen muy similares, tienen significados muy diferentes.

Reconocimiento es un término genérico que no implica por defecto una verificación o identificación de un individuo. Todos los sistemas biométricos realizan reconocimiento para “distinguir de nuevo” una persona que se ha ingresado previamente al sistema.

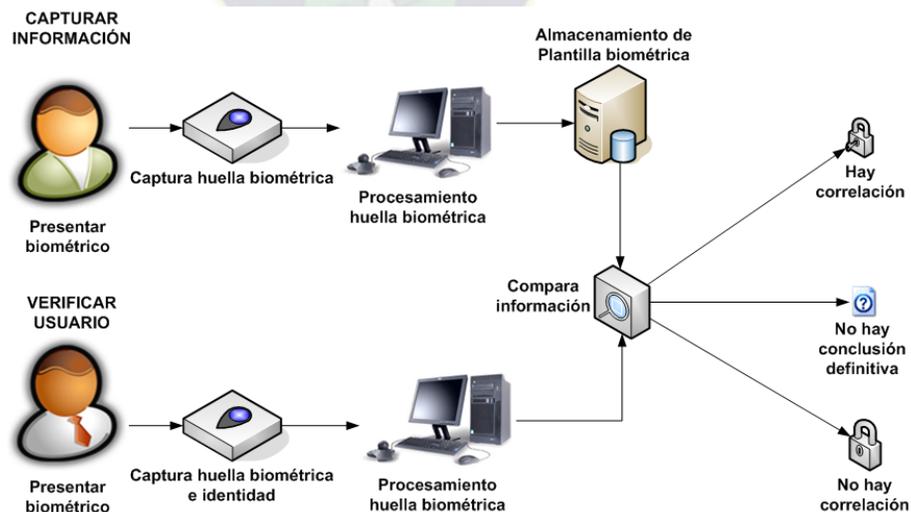
Verificación: Es una tarea de los sistemas biométricos que busca confirmar la identidad de un individuo que la reclama comparando una muestra biométrica con la plantilla biométrica previamente ingresada al sistema.

Identificación: es una tarea donde los sistemas biométricos buscan determinar la identidad de un individuo. El dato biométrico es tomado y comparado contra las plantillas en la base de datos, la identificación puede ser cerrada (si se sabe que la persona existe en la base de datos) o abierta (si no se sabe con certeza si la persona existe en la base de datos), la identificación abierta también es llamada watchlist.

Partiendo de las definiciones anteriores sabemos que hay tres formas para comparar la muestra biométrica, la comparación uno a uno (Verificación), la comparación uno a muchos (Identificación cerrada) y la comparación uno a pocos que es una mezcla de los dos primeros (identificación abierta o watchlist).

Verificación: En el proceso de comparación uno a uno, el usuario presenta su(s) dato(s) biométrico(s) y este se compara con la plantilla biométrica almacenada en una base de datos o en un dispositivo portátil, verificando si hay o no coincidencia para esa identidad en la referencia establecida.

Figura 2.3. Proceso de captura y verificación de usuario



Fuente [Fundamentals of Biometric Authentication Technologies, James I. Wayman]

Identificación cerrada: En el proceso de comparación uno a muchos, el usuario presenta su(s) dato(s) biométrico(s) y el dato biométrico se compara contra la base de datos, donde se sabe que existe, buscando la identidad más probable del usuario.

Identificación abierta: es un proceso híbrido entre la verificación y la identificación cerrada, donde la persona no reclama una identidad específica, entonces se compara contra toda la base de datos para verificar si existe en la base de datos, una vez se verifica que posiblemente existe, dentro de las coincidencias más probables, determina quién es el usuario.

Para la toma de decisiones el resultado de cualquiera de las comparaciones que se hagan puede presentar una de tres posibilidades dependiendo la puntuación que se alcance en la comparación de la plantilla y el dato biométrico y del umbral que se le haya dado al sistema; las tres posibles alternativas son:

- Hay correlación: es decir que al comparar el dato biométrico capturado con la(s) plantilla(s) almacenada(s) la puntuación esta dentro de los umbrales de coincidencia.
- No hay correlación: es decir que al comparar el dato biométrico capturado con la(s) plantilla(s) almacenada(s) la puntuación esta fuera de los umbrales de coincidencia.
- Imposibilidad de alcanzar conclusión definitiva: es decir que hay falta de información para poder hacer una comparación adecuada.

2.2.5. PARÁMETROS BIOMÉTRICOS

La precisión de un sistema biométrico esta determinado por una serie de pruebas, que están divididas en tres categorías tecnología, escenario y operacional y para su evaluación se consideran varios conceptos que se pueden generalizar en dos conceptos la probabilidad de que alguien autorizado sea rechazado y la probabilidad de que alguien no autorizado

sea aceptado, el termino a usar varía, a grandes rasgos, dependiendo el tipo de comparación que se haga y en que categoría se haga la evaluación.

Los términos más comúnmente observados son los siguientes:

- La Tasa de falsa aceptación: (FAR – False Acceptance Rate) Es una estadística que muestra la actuación del biométrico, típicamente cuando opera en la tarea de verificación. En general entre más bajo sea el valor de la tasa de falsa aceptación, más alto es la precisión del sistema biométrico. En esta tasa se muestra el porcentaje de número de veces que el sistema produce una falsa aceptación. Es decir cuando un individuo es identificado como usuario de manera incorrecta. Este valor debe ser lo suficientemente bajo como para que no se impida el ingreso a los usuarios, pero no tanto que permita el ingreso de personal no autorizado. El valor depende de lo sensible del área o sistema a proteger y de la necesidad del usuario. A nivel de fabricantes la mayoría tienen esta tasa entre el 0.0001% y el 0.1%. La tasa dada normalmente asume intentos pasivos del impostor.

$$FAR= PR \times FMR \times (1-FTA).$$

Ec. 2.1

- Tasa de Falso Rechazo (FRR - False Reject Rate): La probabilidad de que un dispositivo rechace una persona autorizada. Comercialmente su valor varía entre el 0.00066% y el 1%.

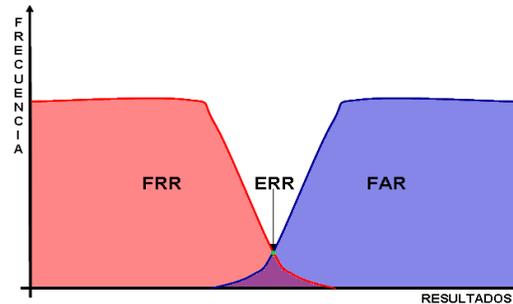
$$FRR=FTA+(1-FTA)\times BER +(1-FTA)\times(1-BER)\times FNMR$$

Ec. 2.2

- El punto de intersección entre la tasa de falsa aceptación y la tasa de falso rechazo se conoce como la tasa de error igual (EER - Equal Error Rate), algunas veces se llama tasa de error cruzada (CER – Crossover Error Rate). Es una estadística que muestra la actuación del biométrico, típicamente cuando opera en la tarea de verificación. En general entre

más bajo sea el valor de la tasa de error igual, más alto es la precisión del sistema biométrico.

Figura 2.4 Definición de la tasa de error igual.

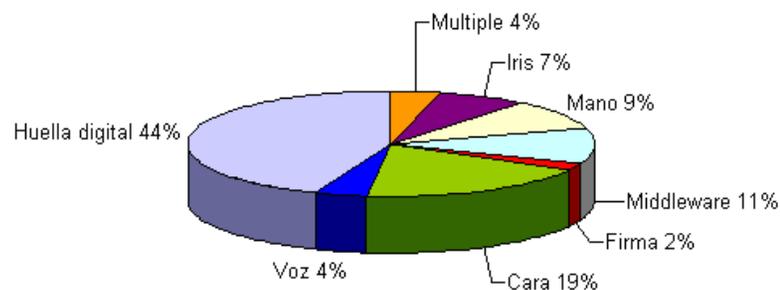


Fuente [Fundamentals of Biometric Authentication Technologies, James I. Wayman]

2.2.6. MODALIDADES BIOMÉTRICAS

Las tecnologías biométricas de mayor uso hoy y con más apoyo por las industrias comerciales son: la huella digital, el reconocimiento facial, la geometría de la mano, el iris, la voz, la firma.

Figura 2.5 Mercado de Biométricos por tecnología



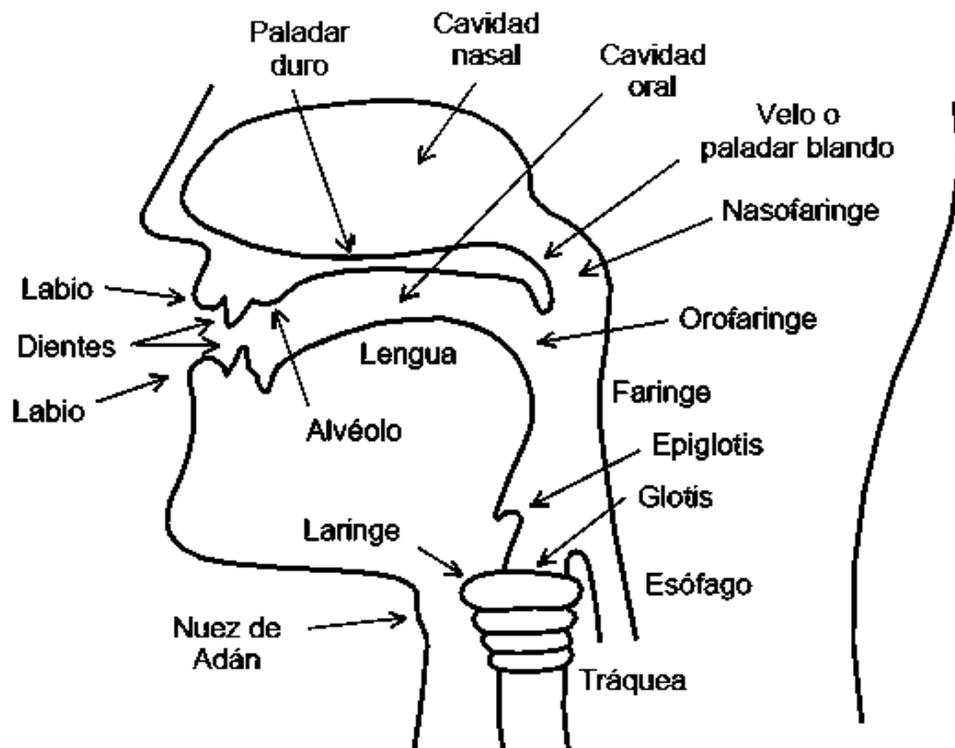
Fuente [Fundamentals of Biometric Authentication Technologies, James I. Wayman]

2.2.7. GENERACIÓN DE LA VOZ

Para determinar las operaciones de un sistema automático de reconocimiento de voz, es fundamental conocer y determinar los mecanismos que han producido un mensaje hablado, para, a continuación poder reproducirlos automáticamente. Es por ello que se van a citar algunos conceptos fundamentales y básicos en el mecanismo de producción del habla, tanto en el órgano físico que soporta dicho mecanismo, como la producción propia del mensaje.

El aparato vocal consta esencialmente de los siguientes órganos que se ilustran a continuación: pulmones, tráquea, laringe y los tractos vocal y nasal.

Figura 2.6 Aparato fonatorio



Fuente [La voz Humana, *Federico Miyara*]

La laringe alberga las cuerdas vocales, que son pequeñas membranas de piel que vibran al pasar el aire entre ellas. El hueco que existe entre ambas se llama glotis. El tracto vocal, de unos 17cm en un adulto, es un tubo que se extiende desde los labios hasta la laringe. El tracto nasal es también un tubo de unos 12cm que va desde los orificios nasales hasta el velo del paladar.

El velo del paladar controla el acoplo entre los dos tractos de la siguiente manera: en sonidos no nasales éste se cierra y el aire pasa por el tracto vocal. En los sonidos nasales el velo permite pasar el aire hacia el tracto nasal, cerrándose el vocal en los labios. En los sonidos nasalizados el aire pasa por ambos tractos.

La función de los pulmones en la producción de habla es de gran importancia. Éstos son los encargados de impulsar el aire hacia el resto de los órganos con la ayuda muscular del diafragma.

Los sonidos de que consta el habla se pueden clasificar básicamente en tres tipos:

- *Sonoros*. Son aquellos sonidos que hacen vibrar las cuerdas vocales. Esta vibración es casi periódica y su espectro es muy rico en armónicos que son múltiplos de la frecuencia de vibración de las cuerdas. A esta frecuencia de vibración de las cuerdas se le llama frecuencia fundamental. La frecuencia fundamental depende de la presión ejercida al pasar el aire por las cuerdas, y de la tensión de éstas.

Sordos. En estos sonidos, que se caracterizan por no provocar la vibración de las cuerdas vocales, se distinguen dos variedades diferentes: fricativos y aspirados.

En los sonidos fricativos se produce un estrechamiento del tracto vocal por el que se hace pasar el aire, lo que proporciona como resultado una excitación de ruido aleatorio.

En los sonidos aspirados la "turbulencia" en el paso de aire se produce en la glotis.

- *Plosivos*. Estos sonidos se producen por la existencia de una obstrucción temporal al paso del aire.

2.2.7.1. TONO

Es la altura musical de la voz. Según el tono, las voces humanas se clasifican en agudas o graves. La escala de registros de altura permite clasificar a las voces masculinas, por lo común, en tres categorías: tenor, barítono y bajo. Existen también tipos de voces intermedias. Desde el punto de vista oratorio, la mejor voz es la del barítono. Timbre: Es el matiz personal de la voz. Es un fenómeno complejo y está determinado por el tono fundamental y los armónicos o tonos secundarios.

2.2.7.2. CANTIDAD

Es la duración del sonido. Según la cantidad, los sonidos pueden ser largos o breves, con toda la gama intermedia de semilargos, semibreves, etc. La cantidad suele depender, en general, de las características de cada idioma, de los hábitos lingüísticos de las regiones o países, de la psicología del habitante, etc.

2.2.7.3. INTENSIDAD

Es la mayor o menor fuerza con que se produce la voz. Hay voces fuertes y voces débiles. En fonética, se denomina acento al conjunto de los anteriores elementos, cuya combinación especial en cada idioma, en cada región de

un mismo idioma y aun en cada individuo, da a ese idioma o habla una característica distintiva.

2.2.8. FONEMAS

El habla se puede ver como una secuencia de unidades básicas de sonido o fonemas. Los fonemas son unidades lingüísticas abstractas y no pueden observarse directamente en la señal de voz. Un mismo fonema se aplica a muchos sonidos ligeramente diferentes llamados alófonos.

Podemos clasificar los fonemas atendiendo a dos criterios: modo de articulación y punto de articulación. En el castellano se definen 24 fonemas que se clasifican en la siguiente tabla de dos entradas (tabla 2.1) atendiendo a los dos criterios enunciados. Así mismo se indica el carácter sonoro (SN) o sordo (SR) del fonema.

Tabla 2.1 Clasificación de los fonemas en castellano

Punto de articulación	Abierto		Labiales				Dentales		Alveolares		Palatales		Velares		Glotales	
			Bilabiales		Labiodentales											
	SN	SR	SN	SR	SN	SR	SN	SR	SN	SR	SN	SR	SN	SR	SN	SR
Plosivas			b	p					d	t			g	k		
Nasales			m						n			ɲ				
Laterales									l			ll				
Fricativas						f				s	y			x		h
Vibr simple									r							
V. Comp.									r							
Africadas												ç				
Vocales	a										e,i		o,u			
Semivocales			w								j					

Fuente [Departamento de Informática. Universidad Carlos III de Madrid, Belén Ruiz Mezcua]

Las vocales en castellano no se suelen clasificar de la manera anterior sino que responden a una clasificación más sencilla atendiendo a la posición de la lengua (anterior, media o posterior) y a la abertura de la boca (cerradas, medio cerradas o abiertas), como se ilustra en la tabla siguiente.

Tabla 2.2 Clasificación de los fonemas vocálicos

Posición de la lengua Abertura de la boca	ANTERIORES	CENTRALES	POSTERIORES
CERRADAS	i		u
MEDIO CERRADAS	e		o
ABIERTAS		a	

Fuente [Departamento de Informática. Universidad Carlos III de Madrid, Belén Ruiz Mezcua]

2.2.9. COARTICULACIÓN Y PROSODIA

El habla es mucho más que la simple secuencia de fonemas. La articulación de un fonema en realidad no es la ideal, sino que viene modificada por la articulación de los sonidos de alrededor. Este fenómeno se conoce por coarticulación, y da lugar a un gran número de variedades alofónicas. La coarticulación es el mecanismo que hace que una secuencia de sonidos tenga al oírla una apariencia natural, y es una de las principales dificultades en la síntesis de voz.

Una secuencia de fonemas coarticulados tampoco hace habla, el habla además de presentar efectos coarticulatorios, presenta variaciones de su frecuencia fundamental, la intensidad y el ritmo a lo largo del tiempo. Entonación, intensidad y ritmo constituyen la prosodia. Entre los ejemplos de la información que proporciona la prosodia se incluye el carácter enunciativo o interrogativo de una frase por la variación de la frecuencia fundamental al final de ésta; la acentuación (incremento de la intensidad y

variación del ritmo) a menudo distingue dos palabras que constan de los mismos sonidos.

2.2.10. MÉTODOS DE ANÁLISIS DE VOZ

El análisis de la señal de voz se puede abordar desde dos puntos de vista: estudiando las características temporales de la voz, o analizando las características espectrales de la voz. Al primero se le llama análisis en el dominio del tiempo, e incluye medidas de parámetros como la tasa de cruces por cero, la autocorrelación, etc. El segundo, llamado análisis en el dominio de la frecuencia, pretende localizar parámetros de la señal de voz atendiendo a la información que provee su espectro. Este último es el que cobra mayor importancia debido a su utilización en los sistemas automáticos del habla.

Aunque la voz cambia a lo largo del tiempo, y por tanto no se puede hablar de estacionariedad, la voz está producida por unos órganos físicos que tienen una cierta inercia mecánica o lo que es lo mismo, no pueden cambiar demasiado rápido. La voz es generada mientras los órganos permanecen en una cierta posición, sí se puede considerar estacionaria, y en su estudio se pueden emplear numerosas técnicas.

Persiguiendo la adquisición de segmentos de voz estacionarios se ha desarrollado el análisis a corto plazo de la voz, que consiste en tomar muestras de voz cada 10 a 30 ms de manera que la voz se pueda considerar estacionaria en este intervalo.

La mayoría de los sistemas de análisis usan este método y realizan el análisis de tramas de voz de 10 a 30 ms.

Es importante conseguir tramas estacionarias de voz para poder realizar su análisis espectral.

2.2.10.1. MUESTREO Y CUANTIFICACIÓN

Muestreo consiste en el proceso de conversión de señales continuas a señales discretas en el tiempo, es un paso para digitalizar una señal analógica.

Este proceso se realiza midiendo la señal en momentos periódicos del tiempo.

2.2.10.1.1. TEOREMA DE NYQUIST

Si $x[n]$ es una secuencia de muestras obtenida a partir de una señal continua en el tiempo $x(t)$, por medio de la relación

$$x[n] = x(nT), \text{ para } -\infty \leq n \leq \infty \quad \text{Ec. 2.3}$$

Donde T es el periodo de muestreo, y su recíproco, $f_s = \frac{1}{T}$ es la frecuencia de muestreo, en muestras por segundo. También podemos expresar la frecuencia de muestreo como $\Omega_s = \frac{2\pi}{T}$ en radianes por segundo.

Entonces el teorema de muestreo de Nyquist está definido como: sea $x(t)$ una señal limitada en banda por:

$$X(j\Omega) = 0 \text{ para } |\Omega| \geq \Omega_N \quad \text{Ec. 2.4}$$

Entonces $x(t)$ está únicamente determinada por sus muestras

$$x[n] = x(nT), \quad n = 0, \pm 1, \pm 2, \pm 3, \dots \text{ si } \Omega_s = \frac{2\pi}{T} \geq 2\Omega_N.$$

La frecuencia Ω_N es comúnmente referida como la frecuencia de Nyquist, y la frecuencia $2\Omega_N$ que tiene que ser excedida por la

frecuencia de muestreo es llamada la razón de Nyquist.

2.2.10.1.2. CUANTIFICACIÓN

En la cuantificación el valor de cada muestra de la señal se representa como un valor elegido de entre un conjunto finito de posibles valores.

Se conoce como error de cuantificación (o ruido), a la diferencia entre la señal de entrada (sin cuantificar) y la señal de salida (ya cuantificada), interesa que el ruido sea lo más bajo posible. Para conseguir esto y según sea la aplicación a desarrollar, se pueden usar distintas técnicas de cuantificación:

- Cuantificación uniforme
- Cuantificación logarítmica
- Cuantificación no uniforme
- Cuantificación vectorial

2.2.10.1.3. CUANTIFICACIÓN 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. Pero son los mas fáciles y menos costosos a implementar.

2.2.10.1.4. CUANTIFICACIÓN LOGARÍTMICA

Para evitar desperdicio de niveles de reconstrucción y de 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 logarítmico antes de la cuantificación. Esta señal comprimida puede ser cuantificada uniformemente. A la salida del sistema la señal pasa por un expansor. A esta técnica se le llama compresión.

2.2.10.1.5. CUANTIFICACIÓN NO UNIFORME

Este cuantificador utiliza la función de la distribución de probabilidad, conociendo esto se puede ajustar los niveles de reconstrucción a la distribución de forma que se minimice el error cuadrático medio.

2.2.10.1.6. CUANTIFICACIÓN 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.

2.2.10.2. ELIMINACIÓN DEL RUIDO

La señal digitalizada es escaneada y las zonas de silencio son removidas por medio del cálculo de energía en corto tiempo. Segmentos de 10ms son analizados para este propósito. En un segmento la energía promedio es menor que un valor umbral proporcional a la energía promedio de la señal entera es descartado.

La siguiente ecuación representa la energía promedio de cada segmento:

$$E_n = \sum_{k=1}^{W_n} |x[k]|^2 w[n-k] \quad \text{ec. 2.5}$$

E_{avg} Representa la energía promedio de la señal entera:

$$E_{avg} = \frac{1}{N} \sum_{k=1}^N |x[k]|^2 \quad \text{ec. 2.6}$$

Se debe considerar que $E_n \geq THRES * E_{avg}$ cada segmento debe ser mayor o igual al promedio de la señal entera multiplicada por un umbral.

2.2.10.3. PRE-ÉNFASIS

En el espectro de los segmentos de voz sonoros, la amplitud decae con la frecuencia debido fundamentalmente a la característica espectral de emisión de las cuerdas vocales. Aunque este efecto se ve compensado por las características de emisión de la boca, la disminución aún es significativa por lo que es necesario compensarla.

Se ha comprobado experimentalmente que esta disminución es de 6dB/octava, lo que hace que cada vez que se dobla la frecuencia, la amplitud se reduzca a 1/16 de la original.

El mecanismo usado para compensar este efecto se denomina pre-énfasis, y consiste en la implementación de un filtro paso alto de pendiente 6db/octava.

En tecnologías digitales el filtrado paso alto más sencillo se realiza mediante la operación

$$y[n] = x[n] - ax[n-1] \quad \text{ec. 2.7}$$

Donde $y[n]$ es la muestra de salida del filtro de pre-énfasis, y $x[n]$ y $x[n-1]$ las muestras actuales y anterior respectivamente. La transformada z de la anterior expresión proporciona una función de transferencia de la forma

$$H(z) = \frac{Y(z)}{X(z)} = 1 - az^{-1} \quad \text{ec. 2.8}$$

Aunque en el caso de segmentos de voz sordos el pre-énfasis no es necesario, por simplicidad, y dado que no afecta negativamente al sistema, se mantiene su uso.

2.2.10.4. 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.

Se realiza el algoritmo en base a la siguiente fórmula:

$$Q_n = \sum_{k=-\infty}^{\infty} x[k]w[n-k] \quad \text{Ec. 2.9}$$

A continuación se muestran los distintos tipos de ventanas que existen:

2.2.10.4.1. VENTANAS DE ANÁLISIS

Las ventanas de análisis más utilizadas se definen para $0 < m \leq N_w$ según:

a) **Ventana rectangular:**

$$w_R(m; N_w) = 1 \quad \text{ec. 2.10}$$

b) **Ventana de Hanning:**

$$w_h(m; N_w) = \frac{1}{2} - \frac{1}{2} \cos(2\pi m / N_w) \quad \text{ec. 2.11}$$

c) Ventana de Hamming

$$w_H(m; N_w) = \frac{27}{50} - \frac{23}{50} \cos(2\pi m / N_w) \quad \text{ec. 2.12}$$

d) Ventana de Bartlett

$$w_B(m; N_w) = \begin{cases} 2m / N_w & \text{si } 0 < m \leq N_w / 2 \\ 2 - 2m / N_w & \text{si } N_w / 2 < m \leq N_w \end{cases} \quad \text{ec. 2.13}$$

e) Ventana de Blackman

$$w_K(m; N_w) = \frac{21}{50} - \frac{1}{2} \cos(2\pi m / N_w) + \frac{2}{25} \cos(4\pi m / N_w) \quad \text{ec. 2.14}$$

Estas ventanas pueden ser caracterizadas por el tamaño de los lóbulos de la magnitud de su espectro de frecuencias. La ventana rectangular posee el lóbulo central con menor ancho de banda pero la magnitud de los lóbulos laterales decae muy lentamente. La ventana de Blackman posee la mínima amplitud en sus lóbulos laterales pero su lóbulo principal tiene un ancho de banda tres veces mayor al de la rectangular [Kuc, 1988]. En el presente proyecto se utiliza la ventana de Hamming ya que este ofrece una posición media entre resolución frecuencial y distorsión armónica en el proceso de ventaneo.

2.2.11. EXTRACCIÓN DE CARACTERÍSTICAS

En el reconocimiento del habla, la señal de voz pre-procesada se ingresa a un nuevo 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.

Hay que tener en cuenta que el número de parámetros debe ser reducido, para no saturar la base de datos, ya que mientras más parámetros tenga la representación menos fiables son los resultados y mas costosa la implementación.

Existen distintos métodos de análisis para la extracción de características, y se concentran en diferentes aspectos representativos. En este caso analizaremos los dos de mayor importancia para el análisis de la voz:

- Análisis de predicción lineal (LPC)
- Análisis cepstral

2.2.11.1. PREDICCIÓN LINEAL

Se trata de una de las técnicas más potentes de análisis de voz, y uno de los métodos más útiles para codificar voz con 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.

Se fundamenta en establecer un modelo de filtro de tipo todo polo, para la fuente de sonido. La principal motivación del modelo todo polo viene dada porque permite describir la función de transferencia de un tubo, que sin perdidas esta formado por diferentes secciones.

El modelo recibe este nombre porque pretende extrapolar el valor de la siguiente muestra de voz $s(n)$ como la suma ponderada de muestras pasadas

$s(n-1), s(n-2), \dots, s(n-K)$:

$$s(n) \approx -\sum_{k=1}^p \alpha_k s(n-k) \quad \text{Ec. 2.15}$$

Incluyendo un término de excitación $Gu(n)$, la ecuación puede escribirse como una igualdad:

$$s(n) = -\sum_{k=1}^p \alpha_k s(n-k) + Gu(n) \quad \text{Ec. 2.16}$$

Siendo α_k los denominados coeficientes de predicción lineal (LPC), y G, la ganancia de excitación. Por otro lado en el dominio Z la ecuación puede escribirse como:

$$S(z) = -\sum_{k=1}^p \alpha_k z^{-k} S(z) + GU(z) \quad \text{Ec. 2.17}$$

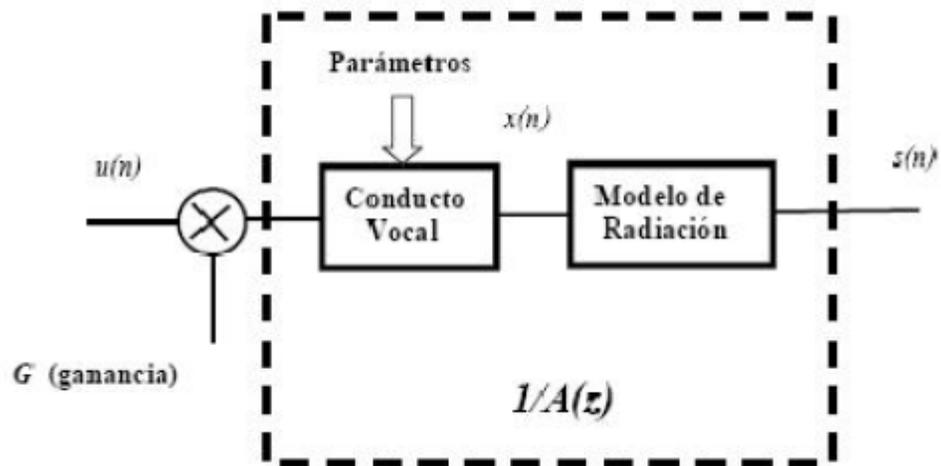
Lo que conduce a una función de transferencia

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 + \sum_{k=1}^p \alpha_k z^{-k}} = \frac{1}{A(z)} \quad \text{Ec. 2.18}$$

La del tipo todo polo. Una interpretación de esta ecuación, que es una versión simplificada, esta dada en la figura 2.7.

$H(z)$ representa la función transferencia de un modelo lineal del conducto vocal + radiación. Los parámetros del filtro digital $H(z)$ son controlados por la señal de voz que está siendo producida y los coeficientes de este filtro son los LPC.

Figura 2.7 Modelo de producción de voz basado en LPC



Fuente [www.pas.deusto.es]

2.2.11.1.1. Estimación de los LPC

Una estima (o predicción) de $s(n)$ basada en p muestras anteriores, puede calcularse como $\hat{s}(n) = -\sum_{k=1}^p \alpha_k s(n-k)$ y el error de estimación (predicción) puede entonces definirse como $\varepsilon(n) = s(n) - \hat{s}(n)$, resultando el error de predicción:

$$\varepsilon(n) = s(n) + \sum_{k=1}^p \alpha_k s(n-k) \quad \text{Ec. 2.19}$$

Los LPC se obtienen minimizando un criterio cuadrático en los errores de predicción, para cada cuadro en que es dividido el segmento de voz.

Suponiendo que en cada cuadro hay $m+1 \gg p$ muestras, y definiendo lo siguiente:

$$\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_p]^T$$

$$\phi^T(n) = [-s(n-1) \ -s(n-2) \ \dots \ -s(n-p)]$$

Ec. 2.20

la ecuación $\hat{s}(n) = -\sum_{k=1}^p \alpha_k s(n-k)$ puede escribirse matricialmente como,

$$\begin{bmatrix} s(n) \\ s(n+1) \\ \vdots \\ s(n+m) \end{bmatrix} = \begin{bmatrix} -s(n-1) & \dots & -s(n-p) \\ -s(n) & \dots & -s(n-p+1) \\ \vdots & \vdots & \vdots \\ -s(n+m-1) & \dots & -s(n+m-1-p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

Ec. 2.21

Donde el vector α son los coeficientes; también se puede escribir lo anterior como,

$$S_m(n) = \phi^T(n) \alpha$$

Ec. 2.22

2.2.11.2. CEPSTRUM

Como se sabe los sonidos de la voz se pueden representar mediante un espectrograma, que indica las componentes frecuenciales de la señal de voz.

Es así entonces como el espectro nos proporciona información acerca de los parámetros del modelo de producción de voz, tanto de la excitación como del filtro que representa el tracto vocal.

Desde el principio de la década de los 70 los sistemas homógrafos han tenido una gran importancia en los sistemas de reconocimiento de voz. Estos sistemas homógrafos son una clase de sistemas no lineales que obedecen a un principio de superposición. De estos los sistemas lineales son un caso especial.

La razón para realizar un procesamiento homomórfico del habla se resume en la figura 2.8.

Figura 2.8 Modelo de la técnica Homomórfica



Fuente [www.pas.deusto.es]

La señal de voz $s(n)$ se descompone en una parte de excitación $e(n)$ y en un filtro lineal $H(e^{i\theta})$, como se mencionó anteriormente. Así, en el dominio de la frecuencia tenemos $S(e^{i\theta}) = H(e^{i\theta})E(e^{i\theta})$

En el dominio logarítmico, por su parte, las dos componentes anteriores pueden separarse empleando técnicas convencionales del procesamiento de señal.

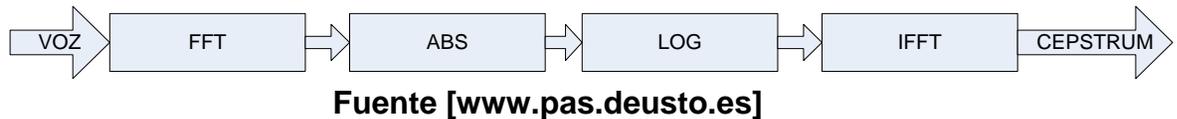
Eso se logra del siguiente modo:

$$\log(|S(e^{j\theta})|) = \log(|H(e^{j\theta})|) + \log(|E(e^{j\theta})|) \quad \text{Ec. 2.23}$$

Para la mayoría de aplicaciones de voz solamente necesitamos la amplitud espectral.

El proceso anterior se puede describir con un diagrama (véase figura 2.9).

Figura 2.9 Modelo de Coeficientes Cepstrales



En la salida de este sistema tenemos entonces:

$$c(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} \log |S_{med}(k)| e^{j \frac{2\pi kn}{N_s}} \quad \text{para } 0 \leq n \leq N_s - 1 \quad \text{Ec. 2.24}$$

En cual caso, el valor $c(n)$ se conoce como coeficientes cepstrales derivados de la transformada de Fourier. N_s es el número de puntos con que se calcula la transformada. Esta ecuación puede ser convenientemente simplificada teniendo en cuenta que el espectro logarítmico es una función real simétrica.

$$c(n) = \frac{2}{N_s} \sum_{k=0}^{N_s} S_{med}(I(k)) \cos\left(\frac{2\pi kn}{N_s}\right) \quad \text{Ec. 2.25}$$

En los cálculos lo habitual es usar solamente los primeros términos $n \leq 20$.

Por otro lado, $I(k)$ representa una función que traduce la posición de un valor en frecuencia al intervalo donde este contenido.

Es posible, a la hora de calcular un coeficiente cepstral, transformar el espectro utilizando bandas definidas según escalas de Mel. En cual caso este tipo de parámetro se conoce como coeficientes cepstrales con frecuencia en escala de Mel (*MFCC*).

Partiendo del análisis de predicción lineal también es posible obtener la expresión de los coeficientes cepstrales asociados:

$$c(0) = \log(1) = 0 \quad \text{Ec. 2.26}$$

$$c(i) = -\alpha(i) - \sum_{j=1}^{i-1} \left(1 - \frac{j}{i}\right) \alpha(j) c(i-j) \quad 1 \leq i \leq N_c \quad \text{Ec. 2.27}$$

En el sistema de reconocimiento de voz en MATLAB existe una función para obtener los coeficientes cepstrales utilizando la FFT. La función utilizada es la *rceps*, que nos proporciona el cepstrum real de la función ingresada, por medio del algoritmo mostrado en la figura 10. O sea que es la implementación del algoritmo mostrado anteriormente. La razón principal para utilizar los coeficientes cepstrales es que tienen la ventaja adicional que uno puede derivar de ellos una serie de parámetros que son invariantes sin importar las distorsiones que puedan ser introducidas por el micrófono o por cualquier sistema de transmisión.

2.2.11.2.1. CARACTERÍSTICAS

Por último, los coeficientes son normalizados para reducir variabilidades espectrales durante largos periodos de tiempo. Los coeficientes son expandidos por medio de una representación polinomial ortogonal durante intervalos de 90ms cada 10ms. Este intervalo es adecuado para preservar información de transición entre fonemas. Solamente los dos primeros coeficientes ortogonales polinomiales son utilizados. Las siguientes ecuaciones se utilizan en el algoritmo:

$$P_{0j} = 1$$

$$P_{1j} = j - 5$$

Ec. 2.28

Los primeros dos coeficientes de la representación ortogonal polinomial son:

$$a = \frac{\sum_j^9 x_j}{9} \quad a = \frac{\sum_j^9 x_j P_{1j}}{\sum_j^9 P_{1j}^2} \quad \text{Ec. 2.29}$$

Los coeficientes a y b representan el promedio, de la función de tiempo de cada coeficiente cepstral en cada segmento respectivamente. Dicha representación es una función del tiempo de los coeficientes cepstrales $x_t(i)$ y los coeficientes polinomiales de primer orden que están representados por $b_t(i)$, donde t es el número de segmento e i es el índice de los coeficientes cepstrales. Como el valor de p es escogido como 10, la representación resultante es una función del tiempo de 20 elementos de características.

2.2.12. MEDIDA DE DISTORSIÓN

Una característica fundamental de los sistemas de reconocimiento de voz es la forma en que los vectores característicos son combinados y comparados con los modelos de referencia.

Para poder realizar estas operaciones es necesario definir una medida de distorsión o distancia entre vectores característicos.

Algunas de las medidas de distorsión más utilizadas son las distancias o métricas inducidas por las normas en espacios.

Por ejemplo, si f_i, f'_i con $i=1,2,\dots,D$ son las componentes de dos vectores característicos f y f' , pueden definirse las siguientes normas L_p .

$$d_1 = \sum_{i=1}^D |f_i - f'_i| \quad \text{Distancia } L_1 \quad \text{ec. 2.30}$$

$$d_2 = \sqrt{\sum_{i=1}^D |f_i - f_i'|^2} \quad \text{Distancia euclidea } L_2 \quad \text{ec. 2.31}$$

Una medida de distorsión muy utilizada para los coeficientes cepstral, que ha probado tener una muy buena performance tanto para reconocimiento de locutor como para reconocimiento de palabra, es la distancia Euclidea ponderada, definida como

$$d_{2w} = \sqrt{\sum_{i=1}^D \left(w_i (c_i - c_i') \right)^2} \quad \text{ec. 2.32}$$

Donde $w_i = \frac{1}{\sigma_i}$

Siendo σ_i^2 una estimación de la varianza del i-ésimo coeficiente cepstral c_i .

Una formulación mas general, que tiene en cuenta la interacción entre coeficientes a través de una matriz de covarianza es la denominada distancia de Mahalanobis, definida como:

$$d_M = \sqrt{(\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu)} \quad \text{ec. 2.33}$$

Donde: Media de los vectores de entrenamiento x_i^e $\mu = \frac{1}{N_e} \sum_{i=1}^{N_e} x_i^e$

Matriz de covarianza de los vectores de entrenamiento x_i^e

$$\Sigma = \frac{1}{N_e} \sum_{i=1}^{N_e} (x_i^e - \mu)(x_i^e - \mu)^T$$

Media de los vectores de reconocimiento x_i^r $\bar{x} = \frac{1}{N_r} \sum_{i=1}^{N_r} x_i^r$

2.2.13. INTERFAZ USB

En un principio teníamos la interfaz serie y paralelo, pero era necesario unificar todos los conectores creando uno más sencillo y de mayores prestaciones. Así nació el USB (Universal Serial Bus) con una velocidad de 12Mb/seg. y como su evolución, USB 2.0, apodado USB de alta velocidad, con velocidades en este momento de hasta 480 Mb/seg., es decir, 40 veces más rápido que las conexiones mediante cables USB 1.1.

USB es una nueva arquitectura de bus o un nuevo tipo de bus desarrollado por un grupo de siete empresas (Compaq, Digital Equipment Corp, IBM PC Co., Intel, Microsoft, NEC y Northern Telecom) que forma parte de los avances plug-and-play y permite instalar periféricos sin tener que abrir la máquina para instalarle hardware, es decir, basta con que se conecte dicho periférico en la parte posterior del computador y listo.

USB Universal Serial Bus es una interfase plug & play entre la PC y ciertos dispositivos tales como teclados, mouses, scanner, impresoras, módems, placas de sonido, cámaras, etc.

Una característica importante es que permite a los dispositivos trabajar a velocidades mayores, en promedio a unos 12 Mbps, esto es más o menos de 3 a 5 veces más rápido que un dispositivo de puerto paralelo y de 20 a 40 veces más rápido que un dispositivo de puerto serial.

2.2.14. FUNCIONAMIENTO DE LA INTERFAZ USB

Trabaja como interfaz para transmisión de datos y distribución de energía, que ha sido introducida en el mercado de PC's y periféricos para mejorar las lentas interfaces serie (RS-232) y paralelo. Esta interfaz de 4 hilos, 12 Mbps y "plug and play", distribuye 5V para alimentación, transmite datos y está siendo adoptada rápidamente por la industria informática.

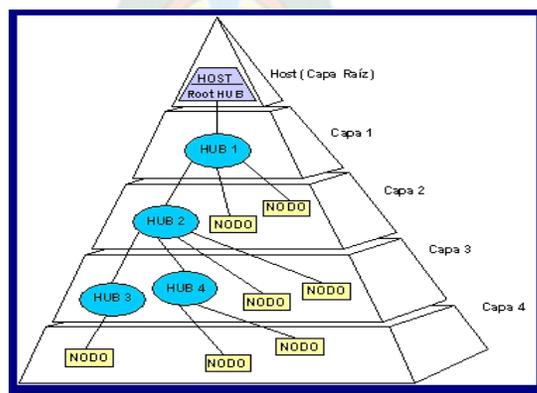
Es un bus basado en el paso de un testigo, semejante a otros buses como los de las redes locales en anillo con paso de testigo. El controlador USB distribuye testigos por el bus. El dispositivo cuya dirección coincide con la que porta el testigo responde aceptando o enviando datos al controlador. Este también gestiona la distribución de energía a los periféricos que lo requieran.

Emplea una topología de estrellas apiladas que permite el funcionamiento simultáneo de 127 dispositivos a la vez. En la raíz o vértice de las capas, está el controlador anfitrión o host que controla todo el tráfico que circula por el bus. Esta topología permite a muchos dispositivos conectarse a un único bus lógico sin que los dispositivos que se encuentran más abajo en la pirámide sufran retardo. A diferencia de otras arquitecturas, USB no es un bus de almacenamiento y envío, de forma que no se produce retardo en el envío de un paquete de datos hacia capas inferiores.

El sistema de bus serie universal USB consta de tres componentes:

- Controlador
- Hubs o Concentradores
- Periféricos

Figura 2.10 Estructura de capas del bus USB



Fuente [Gabriel Echeverria, Claudio Moran]

2.2.15. CONTROLADOR

Reside dentro del PC y es responsable de las comunicaciones entre los periféricos USB y la CPU del PC. Es también responsable de la admisión de los periféricos dentro del bus, tanto si se detecta una conexión como una desconexión. Para cada periférico añadido, el controlador determina su tipo y le asigna una dirección lógica para utilizarla siempre en las comunicaciones con el mismo. Si se producen errores durante la conexión, el controlador lo comunica a la CPU, que, a su vez, lo transmite al usuario. Una vez se ha producido la conexión correctamente, el controlador asigna al periférico los recursos del sistema que éste precise para su funcionamiento.

El controlador también es responsable del control de flujo de datos entre el periférico y la CPU.

2.2.16. CONCENTRADORES Ó HUBS

Son distribuidores inteligentes de datos y alimentación, y hacen posible la conexión a un único puerto USB de 127 dispositivos. De una forma selectiva reparten datos y alimentación hacia sus puertos descendentes y permiten la comunicación hacia su puerto de retorno o ascendente.

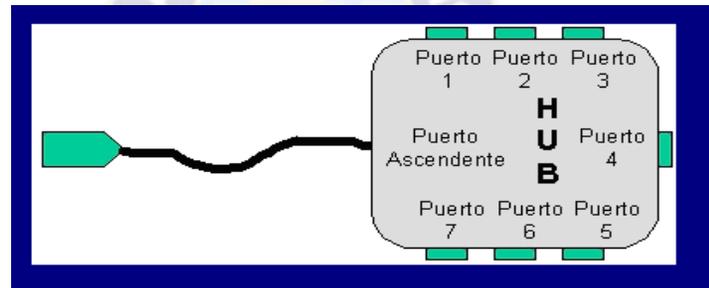
Los concentradores también permiten las comunicaciones desde el periférico hacia el PC, aceptando datos en las 4 puertos descendentes y enviándolos hacia el PC por la puerta de retorno.

Además del controlador, el PC también contiene el concentrador raíz. Este es el primer concentrador de toda la cadena que permite a los datos y a la energía pasar a uno o dos conectores USB del PC, y de allí a los 127 periféricos que, como máximo, puede soportar el sistema. Esto es posible añadiendo concentradores adicionales. Por ejemplo, si el PC tiene una única puerta USB y a ella le conectamos un hub o concentrador de 4 puertos, el PC se queda sin más puertos disponibles. Sin embargo, el hub de 4 puertos

permite realizar 4 conexiones descendentes. Conectando otro hub de 4 puertos a una de las 4 puertos del primero, habremos creado un total de 7 puertos a partir de una puerta del PC. De esta forma, es decir, añadiendo concentradores, el PC puede soportar hasta 127 periféricos USB.

La mayoría de los concentradores se encontrarán incorporados en los periféricos. Por ejemplo, un monitor USB puede contener un concentrador de 7 puertos incluido dentro de su chasis. El monitor utilizará una de ellas para sus datos y control y le quedarán 6 para conectar allí otros periféricos.

Figura 2.11 Esquema de un concentrador



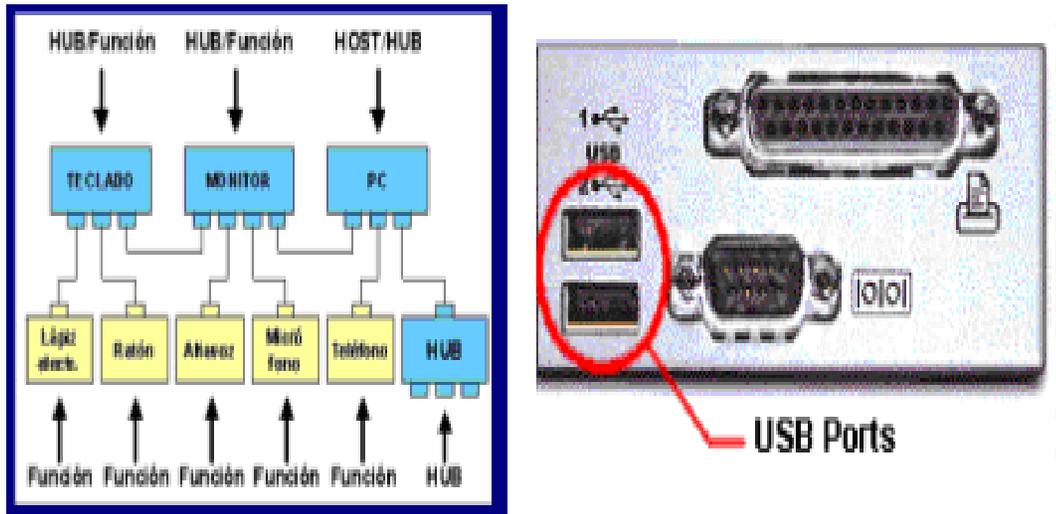
Fuente [Gabriel Echeverria, Claudio Moran]

2.2.17. PERIFÉRICOS

USB soporta periféricos de baja y media velocidad. Empleando dos velocidades para la transmisión de datos de 1.5 y 12 Mbps se consigue una utilización más eficiente de sus recursos. Los periféricos de baja velocidad tales como teclados, ratones, joysticks, y otros periféricos para juegos, no requieren 12 Mbps. Empleando para ellos 1,5 Mbps, se puede dedicar más recursos del sistema a periféricos tales como monitores, impresoras, módems, scanner, equipos de audio, etc.; que precisan de velocidades más altas para transmitir mayor volumen de datos.

En las figuras siguientes se puede ver cómo los hubs proporcionan conectividad a toda una serie de dispositivos periféricos

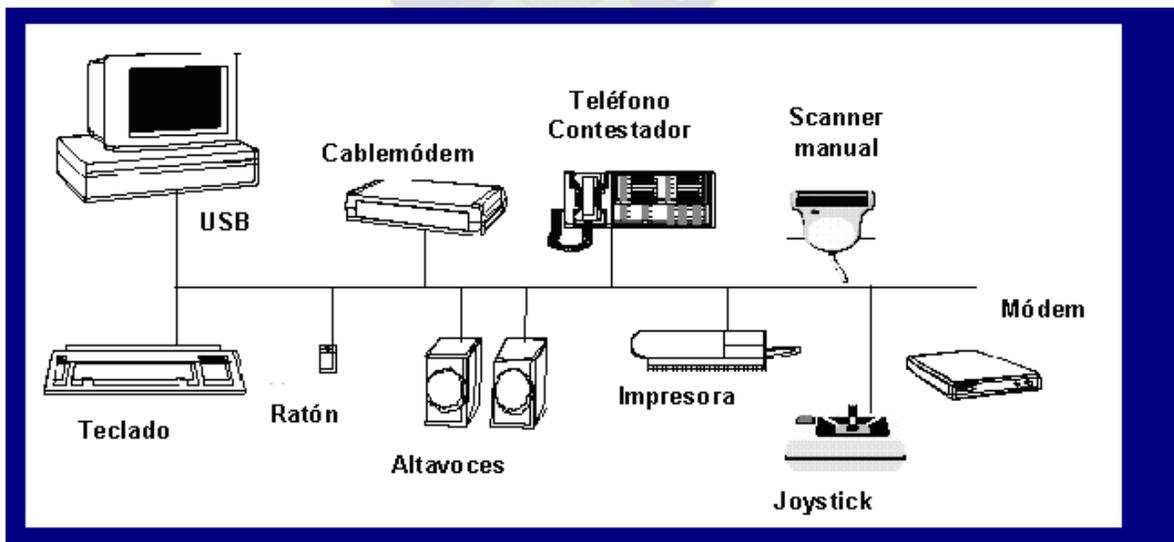
Figura 2.12 Posible esquema de conexiones del bus USB



Fuente [Gabriel Echeverria, Claudio Moran]

En la siguiente figura se observa varios dispositivos USB conectados a una PC. Los dispositivos USB son de diversas aplicaciones y están conectados a un solo bus lógico.

Figura 2.13 Dispositivos USB conectados a un PC

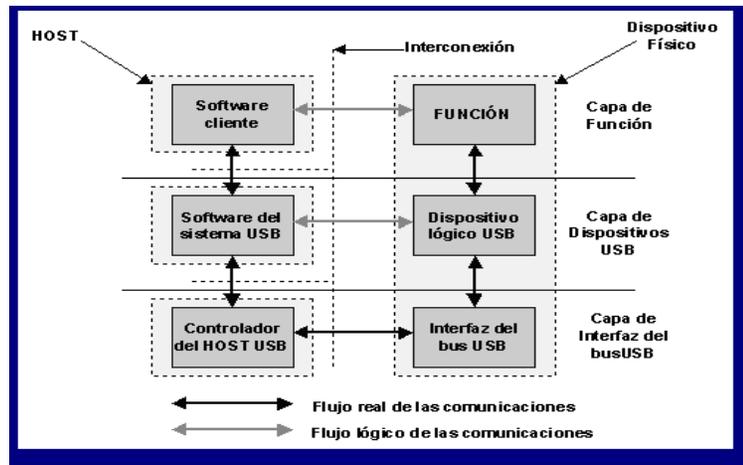


Fuente [Gabriel Echeverria, Claudio Moran]

2.2.18. DIAGRAMA DE CAPAS

En el diagrama de capas de la figura siguiente podemos ver cómo fluye la información entre las diferentes capas a nivel real y a nivel lógico.

Figura 2.14 Diagrama de Capas



Fuente [Gabriel Echeverría, Claudio Moran]

En la figura anterior está materializada la conexión entre el controlador anfitrión o host y un dispositivo o periférico. Este está constituido por hardware al final de un cable USB y realiza alguna función útil para el usuario.

El software cliente se ejecuta en el host y corresponde a un dispositivo USB; se suministra con el sistema operativo o con el dispositivo USB. El software del sistema USB, es el que soporta USB en un determinado sistema operativo y se suministra con el sistema operativo independientemente de los dispositivos USB o del software cliente.

El controlador anfitrión USB está constituido por el hardware y el software que permite a los dispositivos USB ser conectados al anfitrión. Como se muestra en la figura 2.10., la conexión entre un host y un dispositivo requiere la interacción entre las capas. La capa de interfaz de bus USB proporciona la conexión física entre el host y el dispositivo. La capa de dispositivo USB es la

que permite que el software del sistema USB realice operaciones genéricas USB con el dispositivo.

La capa de función proporciona capacidades adicionales al host vía una adecuada capa de software cliente. Las capas de función y dispositivos USB tienen cada una de ellas una visión de la comunicación lógica dentro de su nivel, aunque la comunicación entre ellas se hace realmente por la capa de interfaz de bus USB.

2.2.19. CABLES Y CONECTORES

USB transfiere señales y energía a los periféricos utilizando un cable de 4 hilos, apantallado para transmisiones a 12 Mbps y no apantallado para transmisiones a 1.5 Mbps . En la figura 2.13 se muestra un esquema del cable, con dos conductores para alimentación y los otros dos para señal, debiendo estos últimos ser trenzados o no según la velocidad de transmisión.

Figura 2.15 Esquema del cable para USB



Fuente [Gabriel Echeverria, Claudio Moran]

El calibre de los conductores destinados a alimentación de los periféricos varía desde 20 a 26 AWG, mientras que el de los conductores de señal es de 28 AWG. La longitud máxima de los cables es de 5 metros.

Por lo que respecta a los conectores hay que decir que son del tipo ficha (o conector) y receptáculo, y son de dos tipos: serie A y serie B. Los primeros presentan las cuatro patillas correspondientes a los cuatro conductores alineadas en un plano. El color característico es el blanco sucio y los receptáculos se presentan en cuatro variantes: vertical, en ángulo recto, panel y apilado en ángulo recto así como para montaje pasamuro.

Los conectores de la serie B presentan los contactos distribuidos en dos planos paralelos, dos en cada plano, y se emplean en los dispositivos que deban tener un receptáculo al que poder conectar un cable USB. Por ejemplo impresoras, scanner, y módems.

2.3. MARCO APLICATIVO

Previamente se explicara lo que es paraplejía ya que la Tetraplejía es una sub-clasificación del primero. Empezaremos haciendo énfasis en la médula espinal y como es que esta se daña causando la discapacidad parapléjica.

2.3.1. LA MÉDULA ESPINAL NORMAL

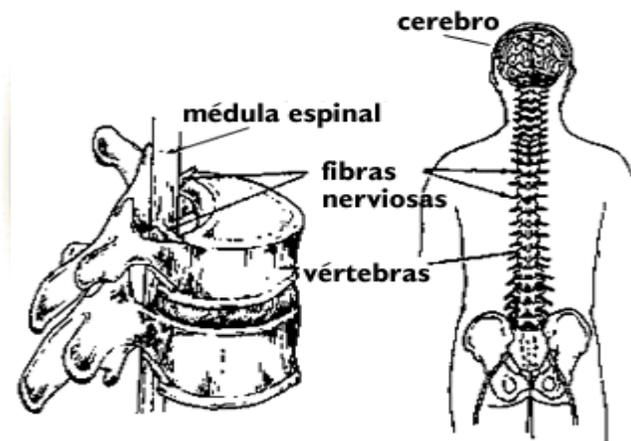
Los nervios son estructuras similares a un cordón compuesto de muchas fibras nerviosas. La médula espinal tiene muchas fibras nerviosas espinales. Las fibras nerviosas transportan mensajes entre el cerebro y las diferentes partes del cuerpo. Los mensajes pueden ser relacionados con el movimiento, diciéndole a alguna parte del cuerpo que se mueva. Otras fibras nerviosas llevan mensajes de sensación o tacto desde el cuerpo hacia el cerebro, como el calor, el frío o el dolor. El cuerpo también tiene un sistema nervioso autónomo. El controla las actividades involuntarias del cuerpo como la presión sanguínea, la temperatura corporal y el sudor.

Estas fibras nerviosas constituyen el sistema de comunicación del cuerpo. La médula espinal puede ser comparada a un cable telefónico. Conecta la oficina principal (el cerebro), con otras oficinas particulares (las partes del cuerpo) por medio de líneas telefónicas (las fibras nerviosas). La médula espinal es el

camino que los mensajes usan para viajar entre el cerebro y las diferentes partes del cuerpo.

Debido a que la médula espinal es una parte vital de nuestro sistema nervioso, está rodeado y protegido por huesos llamados vértebras. Las vértebras, o huesos de la espalda, están colocadas una arriba de la otra y se llama la columna vertebral o la columna espinal. La columna vertebral es el soporte número uno del cuerpo. La médula espinal realmente pasa por en medio de las vértebras. [Vea la Figura 2.16]

Figura 2.16 Medula espinal

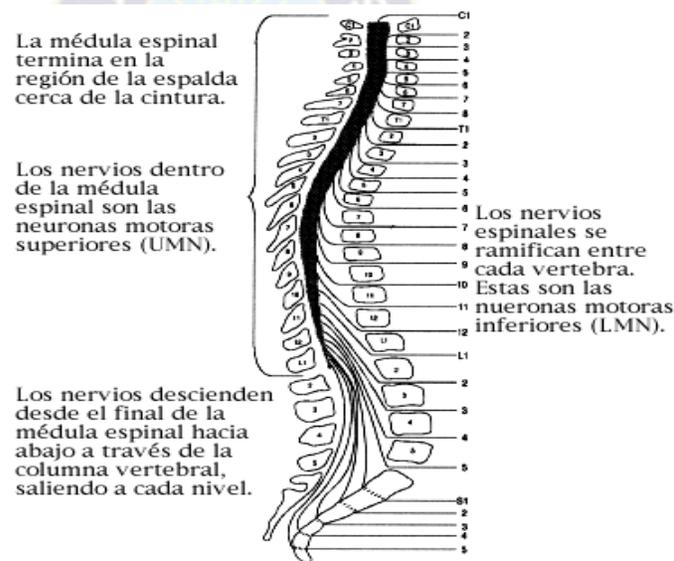


Fuente [www.nccdr.org]

La médula espinal tiene aproximadamente 18 pulgadas de largo. Se extiende desde la base del cerebro. Continúa hacia abajo de la mitad de la espalda, aproximadamente hasta la cintura. El haz de fibras nerviosas que constituyen la médula espinal por si mismo son las neuronas motoras superiores (UMN, por sus siglas en inglés). Los nervios espinales se ramifican desde la médula espinal hacia arriba y hacia abajo del cuello y la espalda. Estos nervios, neuronas motoras inferiores (LMN, por sus siglas en inglés), salen entre cada vértebra y alcanzan todas las partes del cuerpo. La médula espinal termina cerca de la línea de la cintura. Desde este punto, las fibras nerviosas espinales

bajas continúan hacia abajo a través del canal espinal hasta el sacro o coccis. [Vea la Figura 2.16] La columna espinal está dividida en cuatro secciones o partes. La porción superior, es nombrada el área cervical, tiene siete vértebras cervicales. La sección que sigue, la dorsal, incluye el área del pecho y tiene doce vértebras dorsales. La sección baja de la espalda es nombrada el área lumbar. Hay cinco vértebras lumbares. La sección final tiene cinco vértebras sacras y es nombrada el área sacra. Los huesos en la sección sacra, en realidad están fusionados en un solo hueso. [Vea la Figura 2.17]

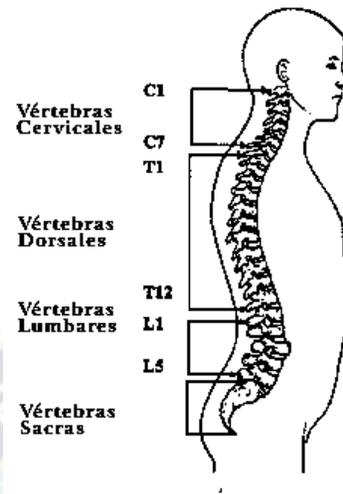
Figura 2.17 División de la médula espinal



Fuente [www.ncddr.org]

Las vértebras son numeradas y nombradas de acuerdo a su situación en la columna espinal. Los nervios espinales son numerados e indican sus vértebras correspondientes.

Figura 2.18 Las vértebras



Fuente [www.ncddr.org]

2.3.2. LA MÉDULA ESPINAL DESPUÉS DE LA LESIÓN

Una lesión de la médula espinal puede ocurrir debido a una lastimadura o daño debido a una enfermedad de la columna vertebral o la médula espinal. En la mayoría de las lesiones de la médula espinal, los huesos de la espalda o vértebras pellizcan la médula espinal. La médula espinal puede inflamarse. La lesión en realidad puede desgarrar el cordón espinal y/o sus fibras nerviosas. Una enfermedad o infección puede causar resultados similares.

Después de una lesión en la médula espinal, todos los nervios por arriba del nivel de la lesión continúan funcionando normalmente. Por abajo del nivel de la lesión, los nervios de la médula espinal no pueden enviar mensajes entre el cerebro y las diferentes partes del cuerpo tal y como lo hacían antes de la lesión.

Cada lesión de la médula espinal es diferente. La lesión en una persona es descrita por su nivel y por su tipo.

2.3.3. LESIÓN PARCIAL Ó TOTAL

El tipo de lesión de la médula espinal es clasificada como parcial o total. La lesión total es como si cortaran todo el servicio telefónico de un edificio. Ningún mensaje puede llegar a las oficinas. Una lesión parcial es como si suspendieran el servicio telefónico en solo algunas oficinas del edificio. Algunas mensajes logran pasar hacia algunas oficinas, mientras otras no. La cantidad y el tipo de mensajes que pueden pasar entre el cerebro y las diferentes partes del cuerpo puede depender de como algunos nervios han sido dañados y otros no.

Algunas personas con una lesión parcial pueden tener mucha sensación pero poco movimiento. Otros pueden tener algo de movimiento y muy poca sensación. Las lesiones parciales espinales, varían de una persona a otra debido a que diferentes fibras nerviosas son dañadas en el cordón espinal de cada persona.

2.3.4. EL NIVEL DE LA LESIÓN

El nivel es el punto más bajo, por debajo del cordón espinal, donde existe una disminución o ausencia de sensación (nivel sensitivo) y movimiento (nivel motor).

Entre más alta sea la lesión del cordón espinal en la columna vertebral, o más cerca esté del cerebro, mayor es la pérdida de la función (sensación y movimiento). Muy pocas partes y sistemas del cuerpo trabajan normalmente con una lesión a un nivel alto.

Por ejemplo, una lesión a nivel de T-8 significa que la persona tiene una disminución o pérdida de sensación y movimiento por debajo del octavo segmento dorsal de la médula espinal. Una persona con una lesión a nivel de C-5 tiene una disminución o pérdida de sensación y movimiento por debajo del quinto segmento cervical de la médula espinal. Alguien con una lesión a nivel

de T-8 podría tener más sensación y movimiento que alguien con una lesión a nivel de C-5. Recuerde que la cantidad de sensación y movimiento también depende de si la lesión es parcial o total.

Se dice que una persona tiene **paraplejía** cuando ha perdido la sensación y no es capaz de mover las partes inferiores ó superiores de su cuerpo. La lesión es en el área dorsal, lumbar o sacra.

Una persona con cuadraplejía, ha perdido movimiento y sensación en ambas partes, superior e inferior de su cuerpo. Esta lesión es en el área cervical.

2.3.5. LA PARAPLEJÍA

La paraplejía se define como la parálisis bilateral simétrica de ambas extremidades inferiores. Según un criterio topográfico céfalo-caudal o, más comúnmente expresado, de arriba abajo, la paraplejía acontece en lesiones bilaterales del córtex motor o área prerrolándica. Entre sus causas cabe destacar, en los niños, los traumatismos obstétricos, lesiones o infecciones cerebrales perinatales que pueden dejar como secuela la paraplejía infantil, como forma de parálisis cerebral del niño

La palabra paraplejía la hemos obtenido directamente del griego. Hipócrates usó ya el término *paraphgia* (*parapleguía*) para referirse a la **parálisis** parcial o ligera, en contraposición a *apoplexia* (*apoplexía*), con la que se referían a la parálisis total. Estos términos tenían para los griegos además otro significado, relativo a las facultades mentales: *apoplexia* (*apoplexía*) significaba tanto apoplejía como locura; y el adjetivo *apoplektos* (*apóplectos*) calificaba al que tiene el espíritu "tocado", de donde derivaban hacia el que ha perdido la razón, el estúpido; y en el mismo plano estaba el lisiado, el impedido; este adjetivo lo usaban los médicos para referirse a todo aquello que tenía que ver con la apoplejía, con la parálisis. Con el otro compuesto ocurre lo mismo: frente a la forma parece que exclusivamente médica *paraphgia* (*parapleguía*), estaba otra más común: *paraplexia* (*paraplexía*; *la diferencia está en la alternancia*

g/x), que significaba ataque de locura, demencia. Y el adjetivo derivado de ambas formas, *paraplhktoV* (*paráplektos*), significa tanto afectado de demencia, como de parálisis ligera o parcial. Y respecto a la hemiplejía, si bien no formaron los griegos este nombre, sí usaron en cambio su adjetivo *hmiplhx* (*hemipléx*) con el significado de medio herido, medio lisiado. El origen lo tenemos, pues, en el verbo *plhssō* (*plésso*), que significa golpear, chocar (no encierra idea de hostilidad), de donde derivará la idea de parálisis, e incluso llegó a significar la parálisis específica de ambas piernas.

Los diccionarios definen la **paraplejía** como la parálisis más o menos completa de partes simétricas: de ambos miembros superiores (paraplejía superior, cervical o braquial –de los brazos-), o de los miembros inferiores (paraplejía inferior o crural –de las piernas-); o de unos y otros (paraplejía braquiocrural, más comúnmente llamada tetraplejía). Pero mayormente se usa este término para denominar la parálisis de las dos extremidades inferiores. Los orígenes de las paraplejías son múltiples, aunque el más divulgado es el de los accidentes de tráfico. En los animales, las causas y la evolución de las paraplejías son muy parecidas a las que se dan en el hombre, por lo que las investigaciones que se hacen en aquéllos son de total aplicación al hombre. Esto ha abierto recientemente enormes posibilidades, al haber conseguido mediante la cirugía intervenir en los mismos centros motores del cerebro y restaurarlos.

En resumen la Paraplejía es una enfermedad permanente y no progresiva en la que hay pérdida de sensibilidad en los miembros paralizados y otros efectos secundarios como:

- Espasmos.
- Dolor.
- Pérdida de control de la función intestinal.
- Pérdida de control de la vejiga de la orina.

Figura 2.19 Persona parapléjica



Fuente [<http://salud.discapnet.es>]

2.3.6. CUADRAPLEJÍA

Además de los síntomas de la paraplejía se suman la debilitación de manos y brazos. Es decir estas personas carecen de movimiento en sus extremidades superiores e inferiores.

Muchas personas con lesión de la médula espinal todavía tienen cierta sensación en las partes paralizadas de su cuerpo, sensación de tacto o dolor.

La enfermedad se acompaña con frecuencia de otras manifestaciones secundarias como úlceras de decúbito, infecciones de vejiga, espasmos musculares y dolor. La fertilidad suele verse afectada en el varón pero no en la mujer.

Como se vera las personas con discapacidad cuadrupléjica en sus distintas clasificaciones, son personas que tienen muchos problemas físicos y necesitan de la tecnología para poder llevar su vida diaria de una manera mas normal, el presente proyecto pretende coadyuvar a mejorar la calidad de vida de estas personas y poderlos ayudar ha hacer de su vida lo mas normal posible.



CAPÍTULO III

INGENIERÍA DEL PROYECTO

CAPÍTULO 3. INGENIERÍA DEL PROYECTO

3.1. INTRODUCCIÓN

En el presente capítulo se desarrolla los procesos del proyecto, se muestra el esquema general del proyecto, la interfaz hardware, además los diagramas de flujo del software de reconocimiento de voz.

Se analiza paso a paso el desarrollo del reconocimiento de voz, las entradas y los tipos de resultados que se obtendrán a partir de los mismos, la grabación de comandos de voz a través de un micrófono, el procesamiento de la señal en el computador y la salida esperada a través de una interfaz USB.

Los requerimientos analizados en este capítulo están basados y conformes con el estándar IEEE 830, el cual busca una descripción comprensiva de las delimitaciones de la arquitectura para así determinar su alcance.

El análisis de los anteriores puntos nos lleva al desarrollo del prototipo cumpliendo así con el objetivo principal del presente proyecto.

3.2. REQUERIMIENTOS

3.2.1. DESCRIPCIÓN GENERAL

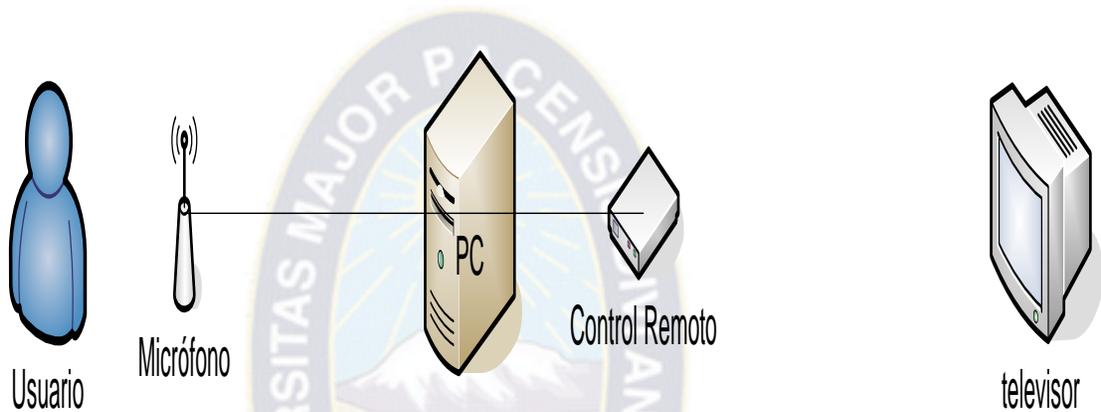
3.2.1.1. PERSPECTIVA DEL PRODUCTO

El sistema de control para un equipo de televisión basado en el reconocimiento de patrones de voz destinado a personas cuádruplégicas es un sistema independiente, cuyo objetivo principal es Desarrollar e implementar un sistema de control para un televisor aplicando reconocimiento de voz para así mejorar la calidad de vida de los discapacitados cuádruplégicos.

3.2.1.2. FUNCIONALIDAD DEL PRODUCTO

El presente proyecto tiene las siguientes funcionalidades: grabar comandos de voz, extraer las características de la voz, reconocimiento de patrones de voz, y cuya finalidad es controlar el equipo de televisión por comandos de voz a través de un puerto USB. Esto se resume en siguiente esquema:

Figura 3.1 Esquema general del proyecto



Fuente [Elaboración propia]

3.2.1.3. CARACTERÍSTICAS DE LOS USUARIOS

A continuación se detalla las características de los usuarios finales:

Tabla 3.1 Características de los usuarios

Tipo de usuario	Discapacitado cuadrapléjico
Formación	No se requiere que tenga algún tipo de formación.
Habilidades	Que posea el habla en buena condición.
Actividades	Que en sus actividades rutinarias utilice el equipo de televisión.

Fuente [Elaboración propia]

3.2.1.4. RESTRICCIONES

- El sistema de control para un equipo de televisión basado en el reconocimiento de patrones de voz destinado a personas cuádruplégicas, solo controlará los siguientes eventos en el equipo de televisión: encendido, apagado, aumentar volumen, disminuir volumen, siguiente canal, anterior canal.
- El sistema se limitará al reconocimiento de comandos de voz y no así de usuarios.
- El sistema se limitará a controlar un equipo de televisión en específico, por ser este el primer prototipo.

3.2.1.5. SUPOSICIONES Y DEPENDENCIAS

Para que el sistema sea lo más eficientemente posible cada nuevo usuario deberá grabar los comandos de voz para que el sistema procese esta señal y pueda extraer las características de su voz y así disminuir el porcentaje de error.

3.2.1.6. EVOLUCIÓN PREVISIBLE DEL SISTEMA

A futuro se pretende aumentar el número de comandos de voz, hacer el reconocimiento de locutores y así hacer que la voz sea una llave biométrica.

Investigar la implementación del mismo sistema de control basada en la tecnología de los microcontroladores para así abaratar los costos de hardware.

3.2.2. REQUISITOS ESPECÍFICOS

El sistema deberá ser capaz de grabar los comandos de voz en tiempo real, para posteriormente proceder a su procesamiento específicamente la extracción de las características. En la siguiente tabla se resume este requisito:

Tabla 3.2 Requisito Especifico 1

Número de requisito	RF1
Nombre de requisito	Grabación de comandos de voz
Tipo	Requisito
Prioridad del requisito	Alta/Esencial

Fuente [Elaboración propia]

El sistema debe ser capaz de reconocer comandos de voz, pronunciados en tiempo real por el usuario final y asociar cada comando de voz con una salida (bit) en un puerto USB. En la siguiente tabla se resume este requisito:

Tabla 3.3 requisito Especifico 2

Número de requisito	RF2
Nombre de requisito	Reconocimiento
Tipo	Requisito
Prioridad del requisito	Alta/Esencial

Fuente [Elaboración propia]

El usuario debe grabar y cambiar los comandos de voz de la base de datos para que el sistema pueda obtener las características de su voz y así hacer que el sistema sea más eficiente. En la siguiente tabla se resume este requisito:

Tabla 3.4 Requisito Especifico 3

Número de requisito	RF3
Nombre de requisito	Modificación
Tipo	Requisito
Prioridad del requisito	Media/Deseado

Fuente [Elaboración propia]

3.2.3. REQUISITOS COMUNES DE LOS INTERFACES

Las entradas al sistema son los comandos de voz en tiempo real y las salidas son el control del equipo de televisión que está específicamente ligado a cada comando de voz.

3.2.3.1. INTERFACES DE USUARIO

La interfaz es una pantalla que sea lo más amigablemente posible para el usuario final, en la que se muestra la grafica del comando de voz en el tiempo y el espectro de frecuencias de la misma. También a cada comando que pronuncie el usuario, el sistema especifica que hace dicho comando ó si es que ha existido algún error.

3.2.3.2. INTERFACES DE HARDWARE

La interfaz entre el PC y el equipo de televisión se realiza mediante un puerto USB asignando a cada comando de voz un BIT específico en dicho puerto y modificando el control remoto que posee dicho televisor. Para la comunicación USB se hace uso de un PIC de gama alta el cual es el PIC 18F4550 que posee un modulo de comunicación USB.

3.2.3.3. INTERFACES DE SOFTWARE

Siendo que la señal de voz es un vector con información de la energía eléctrica que posee la misma en determinados intervalos de tiempo, con estas características se escogió el software para la implementación del sistema específicamente del software de reconocimiento de voz que será MATLAB 7.10 ya que este software se especializa en cálculos numéricos con vectores y matrices.

3.2.4. REQUISITOS FUNCIONALES

El software deberá estar en espera de que se pronuncie un comando de voz, luego extraer las características de la señal de voz, lo cual incluye la eliminación del ruido, el filtrado de pre-énfasis, la segmentación, la obtención de los coeficientes ceptrales, y finalmente la extracción de características para luego hacer una medida de distancia con los parámetros almacenados en la base de datos y así identificar el comando pronunciado.

3.2.5. REQUISITOS DEL TELEVISOR

Los requisitos del televisor son mínimos, este solo deberá poseer su correspondiente control remoto, ya que se modificará el control remoto para hacer que el mismo sea operado por comandos de voz y así alcanzar el objetivo principal del presente proyecto.

Debemos recalcar que el sistema puede ser implementado con cualquier televisor que cumpla con el requisito anterior, sin importar marcas, tamaños, etc.

3.2.6. REQUISITOS NO FUNCIONALES

3.2.6.1. REQUISITOS DE RENDIMIENTO

El sistema responderá a los comandos especificados anteriormente, reconociendo los comandos de voz con un mínimo de error. Siendo que este es el primer prototipo se experimentará con pocos comandos de voz, para experimentar el tiempo de respuesta que dependiendo de la PC está entre los 1 a 3 segundos. En todo momento el sistema deberá estar atento de que se pronuncie algún comando para luego procesar el mismo y dar la respuesta esperada.

3.2.6.2. FIABILIDAD

El sistema deberá tener un mínimo de error y un tiempo de respuesta de entre 1 a 3 segundos. Haciendo así que el sistema sea lo mas fiable posible.

3.2.6.3. DISPONIBILIDAD

El sistema de control una vez iniciada la sesión deberá estar disponible en todo momento es mas deberá estar a la espera de la pronunciación de un comando de voz para procesarlo y ejecutarlo.

3.2.6.4. MANTENIBILIDAD

El mantenimiento puede ser realizado por el usuario que tenga conocimientos básicos de computación. En caso de no ser así se puede comunicar con el autor del mismo o un técnico autorizado. La mantenibilidad en este sistema se refiere básicamente a cambiar la base de datos, modificar los comandos a pronunciar, ampliar el número de comandos, y algunas mejoras o alteraciones que el usuario final vea conveniente.

3.2.6.5. PORTABILIDAD

Para poder ejecutar el software de reconocimiento de voz que es parte del sistema total en un determinado PC previamente se deberá tener instalado MATLAB en sus versiones 7 para adelante. MATLAB funciona en plataformas Windows y Linux debido a esto también lo hará el sistema de control de equipos electrónicos para discapacitados parapléjicos aplicando reconocimiento de voz. Además como la comunicación entre software y hardware que utiliza el presente proyecto es una interfaz USB es posible utilizar cualquier PC previa instalación de los drivers correspondientes.

3.3. ANÁLISIS Y DISEÑO DEL SOFTWARE DE RECONOCIMIENTO DE VOZ

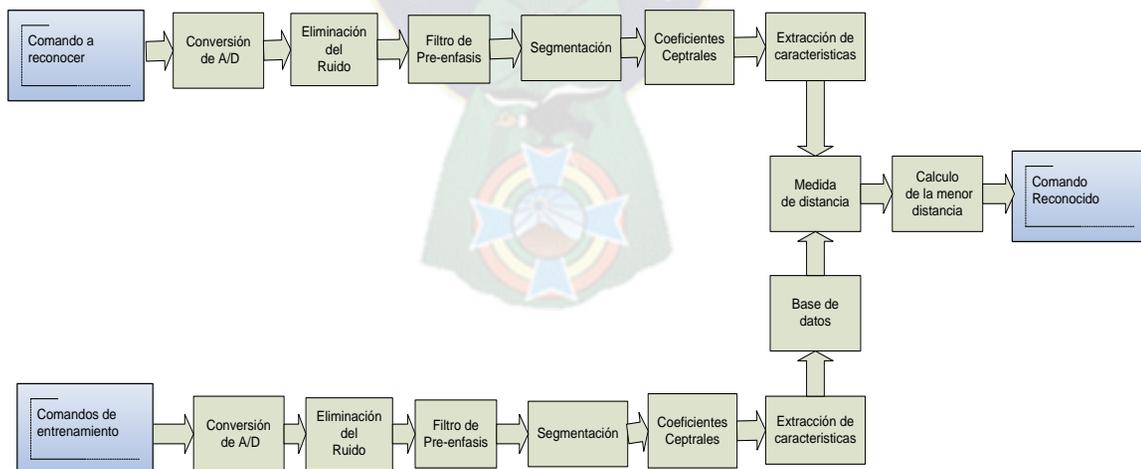
El Procesado de voz es el estudio de las señales de voz y las técnicas de procesado de estas señales. Las señales se digitalizan con el propósito de manipular su información, lo cual es llamado procesamiento digital de voz.

El procesamiento digital de voz se puede dividir en varias categorías, la de nuestro interés es el reconocimiento de voz.

En el reconocimiento de voz el problema radica en identificar las palabras pronunciadas, sin importar el hablante. Bajo este esquema, se pre-procesan las señales de voz, se obtienen las características, y lo que se trata al final es capturar las similitudes entre las palabras habladas.

El algoritmo de reconocimiento de voz se puede resumir en el siguiente diagrama en bloques.

Figura 3.2 Esquema del algoritmo de reconocimiento de voz



Fuente [Elaboración propia]

3.3.1. OBTENCIÓN DE INFORMACIÓN

La obtención de la información, específicamente de las señales de voz a ser reconocidas y las de la base de datos se la obtienen mediante un micrófono.

El micrófono es un transductor electroacústico. Su función es la de transformar (traducir) la presión acústica ejercida sobre su capsula por las ondas sonoras en energía eléctrica.

El audio es un fenómeno analógico. 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 esta 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 esta 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 estereo), pero para las aplicaciones de reconocimiento de voz un canal mono es suficiente. La mayoría de aplicaciones vienen con valores pre-determinados, para desarrollo del código se debería de cambiar los parámetros para ver lo que mejor funciona en el algoritmo.

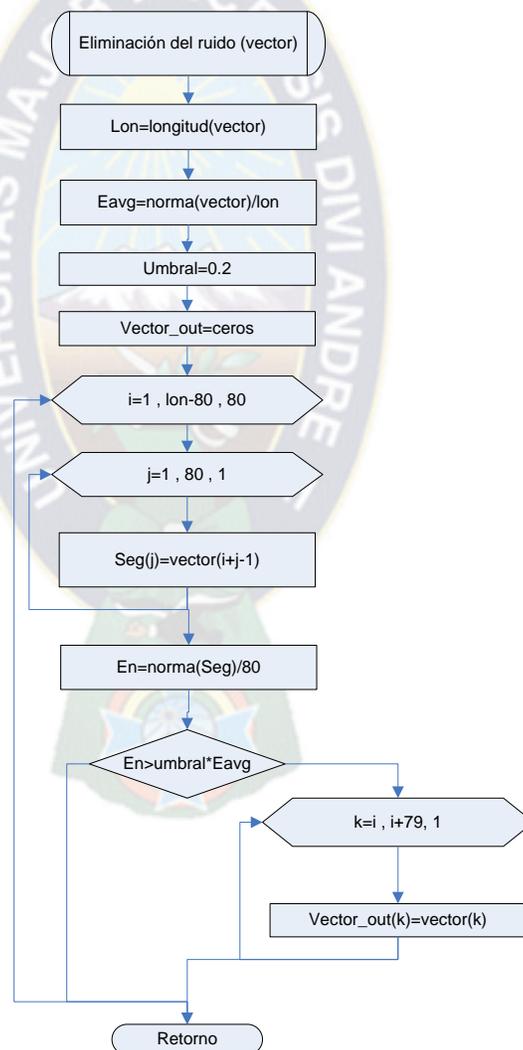
Dado a que el habla es relativamente de bajas frecuencias (entre 100Hz-8kHz), una frecuencia de muestreo de 16000 muestras/seg. provee una mayor exactitud en la adquisición de la información, esta es la frecuencia de nyquist.

3.3.2. ELIMINACIÓN DEL RUIDO

La eliminación del ruido es diseñada con base teórica en el punto 2.2.10.2

Se debe considerar que $E_n \geq THRES * E_{avg}$ cada segmento debe ser mayor o igual al promedio de la señal entera multiplicada por un umbral. El valor umbral escogido es THRES=0.2.

Figura 3.3 Diagrama de flujo de la función eliminación del ruido

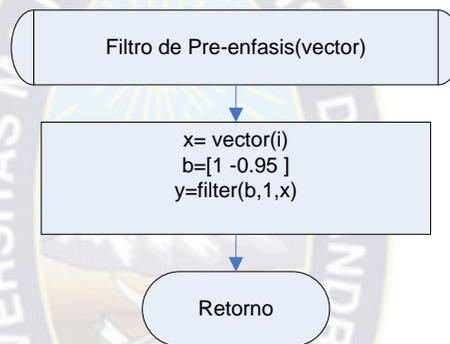


Fuente [Elaboración propia]

3.3.3. FILTRO DE 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 esta contenida en las frecuencias bajas, en segundo remueve la componente DC de la señal, aplanando espectralmente la señal. El siguiente filtro fue diseñado con base referencial al punto 2.2.10.3 del marco teórico.

Figura 3.4 Diagrama en bloques de un filtro



Fuente [Elaboración propia]

3.3.4. 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.

Para el proceso una ventana de Hamming de 30ms es aplicada a la señal de voz, enfatizada previamente con el filtro de pre-énfasis. Con un desplazamiento típico 10ms entre cada ventaneo.

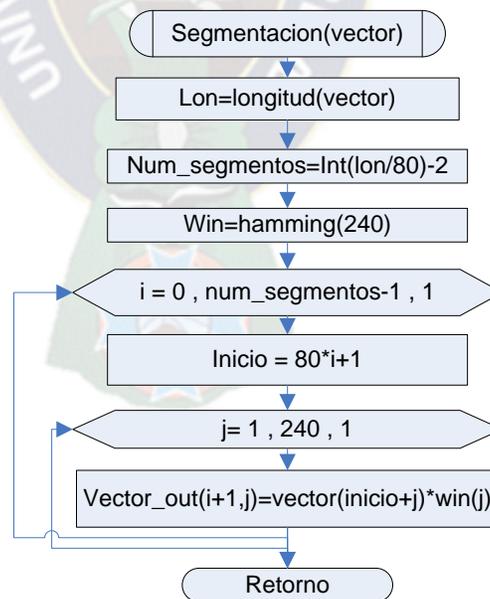
La segmentación se realiza con base teórica en el punto 2.2.10.4 del capítulo II.

Tabla 3.5 Valores de parámetros

Parámetro	Valor
N número de muestras en el segmento de análisis.	240 (30ms)
M número de muestras entre cada segmento	80 (10ms)
p LPC orden de análisis.	10
Q dimensión del vector cepstral derivado del LPC	15

Fuente [Elaboración propia]

Figura 3.6 Diagrama de flujo de la función segmentación

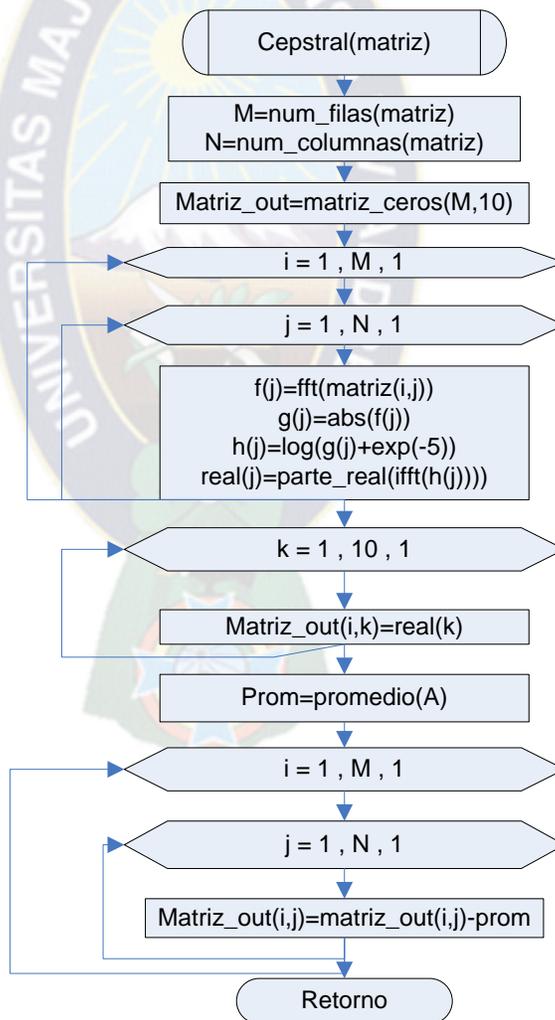


Fuente [Elaboración propia]

3.3.5. COEFICIENTES CEPSTRALES

Los sonidos de la voz se pueden representar mediante un espectrograma, que indica las componentes frecuenciales de la señal de voz. Es así como el espectro nos proporciona información acerca de los parámetros del modelo de producción de voz, tanto de la excitación como del filtro que representa el tracto vocal. El mismo fue diseñado con base teórica respecto del punto 2.2.11.2 del capítulo II.

Figura 3.7 Diagrama de flujo de la extracción de coeficientes cepstrales

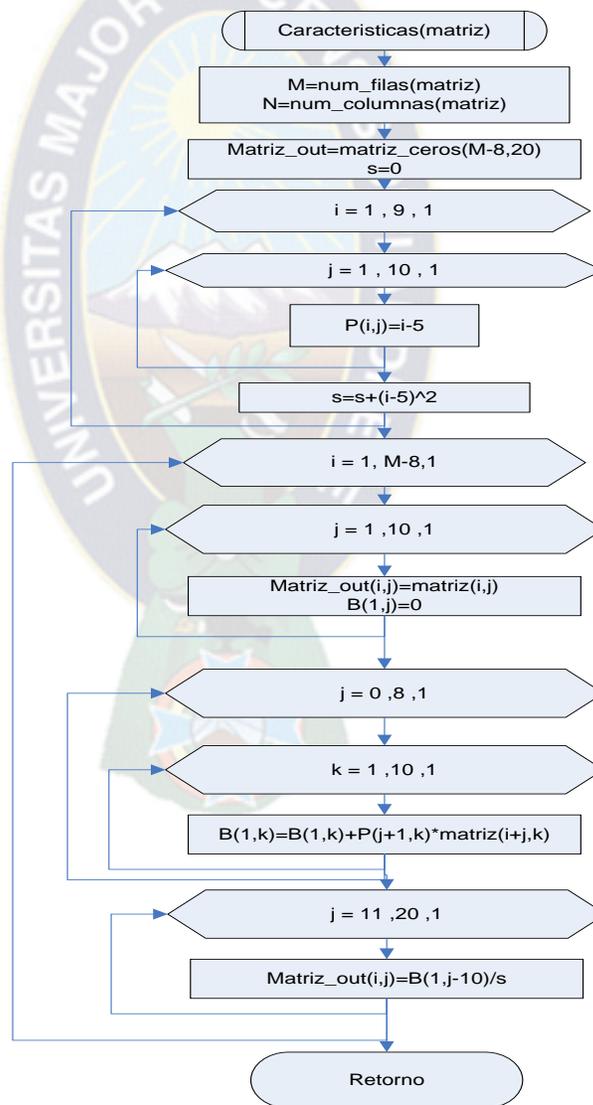


Fuente [Elaboración propia]

3.3.6. EXTRACCIÓN DE CARACTERÍSTICAS

Los coeficientes son normalizados para reducir variabilidades espectrales durante largos periodos de tiempo. Los coeficientes son expandidos por medio de una representación polinomial ortogonal durante intervalos de tiempo de 10ms. Este intervalo es adecuado para preservar información de transición entre fonemas. La siguiente figura muestra el algoritmo de este proceso.

Figura 3.8 Diagrama de flujo de la extracción de características

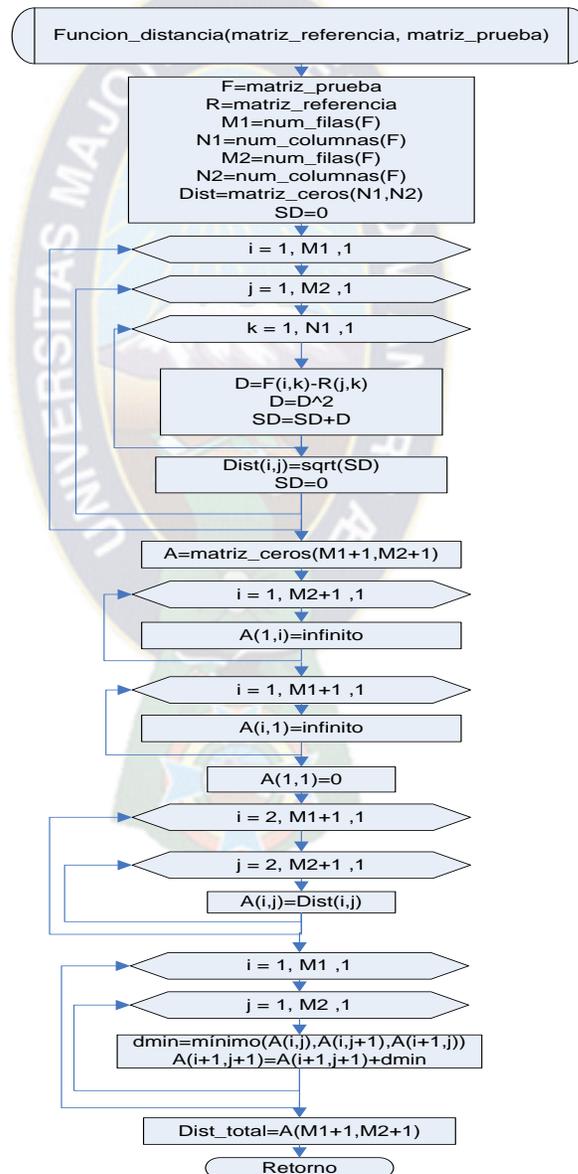


Fuente [Elaboración propia]

3.3.7. MEDIDA DE DISTANCIA

En este paso se calcula la distancia de una muestra respecto de los patrones almacenados en la base datos, el siguiente diagrama de flujo fue realizado en función de las ecuaciones del marco teórico específicamente del punto 2.2.12.

Figura 3.9 Diagrama de flujo de la medida de distancia

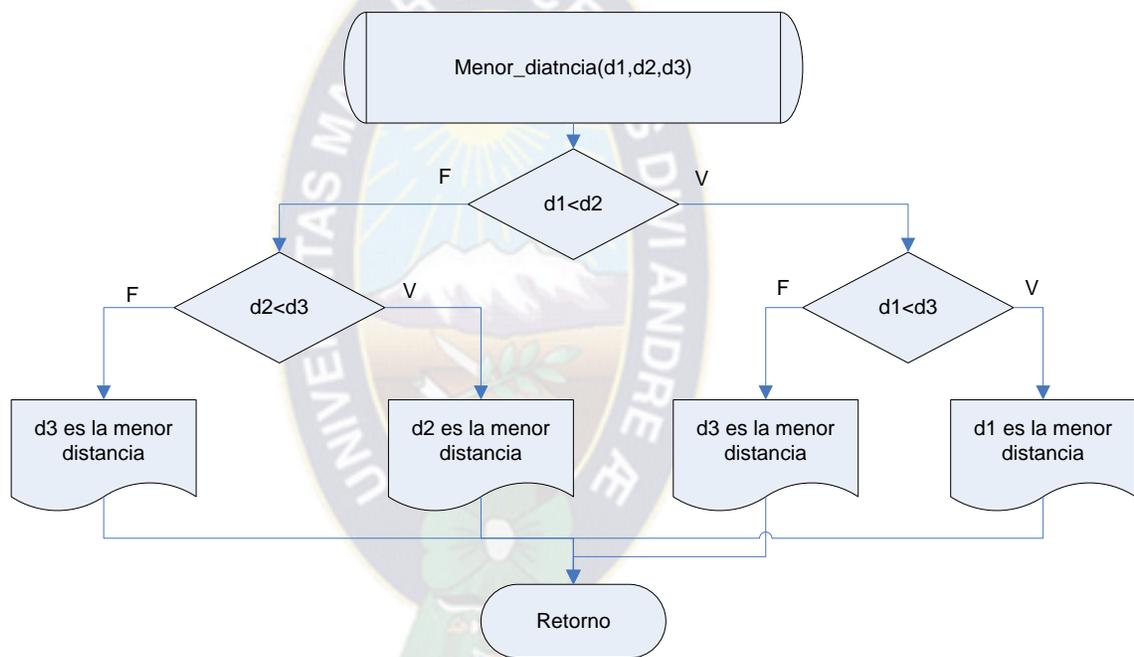


Fuente [Elaboración propia]

3.3.8. CALCULO DE LA MENOR DISTANCIA

En esta etapa se hace el calculo de la menor distancia de un conjunto de distancias, cada distancia esta asociada a un comando de voz a reconocer. Por ejemplo: si se tienen un conjunto de tres distancias provenientes del cálculo del anterior punto, el diagrama de flujo para determinar la menor de las distancias es el siguiente.

Figura 3.10 Diagrama de flujo del calculo de la menor distancia



Fuente [Elaboración propia]

Este diagrama variara dependiendo de cuantos comandos estén almacenados en la base de datos.



CAPÍTULO IV
IMPLEMENTACIÓN Y PRUEBAS DEL SISTEMA

CAPÍTULO 4. IMPLEMENTACIÓN Y PRUEBAS DEL SISTEMA

4.1. INTRODUCCIÓN

En este capítulo se hace énfasis en la implementación y pruebas del prototipo implementado, los algoritmos y procedimientos concernientes al control de un equipo de televisión basado en el reconocimiento patrones de voz para discapacitados cuádruplécicos. Para lo mismo se utilizan los diagramas UML para el modelado del sistema y la mejor comprensión del mismo.

También se muestran las pantallas que se implementaron en el proyecto, la interfaz hardware constituyéndose en un punto fundamental para lograr los objetivos trazados en el capítulo I.

4.2. ANÁLISIS Y DISEÑO DEL PROTOTIPO

Para poder realizar un buen análisis y diseño utilizaremos varias herramientas que nos permitan comprender de la forma mas clara posible los procesos del presente proyecto y así cumplir con los objetivos inicialmente trazados.

4.2.1. MODELO UML

El Lenguaje Unificado de Modelado prescribe un conjunto de notaciones y diagramas estándar para modelar sistemas orientados a objetos, y describe la semántica esencial de lo que estos diagramas y símbolos significan.

UML se puede usar para modelar distintos tipos de sistemas: sistemas de software, hardware, y organizaciones del mundo real.

Los modelos de UML que se utilizaran en específico son:

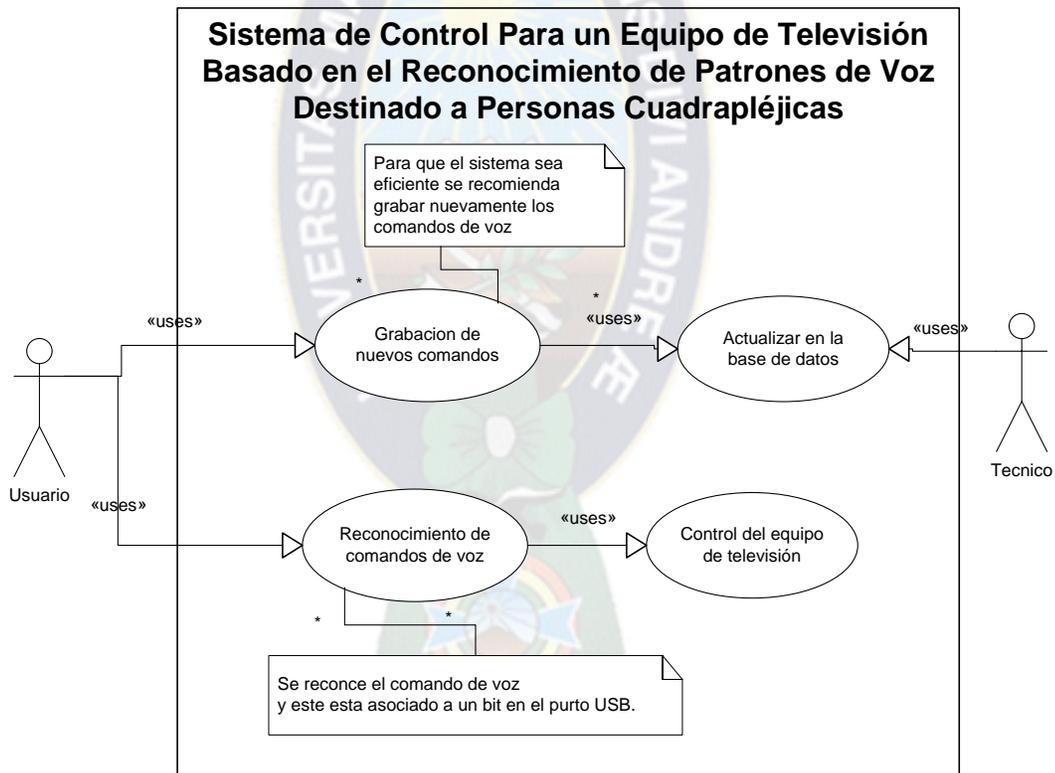
- Diagrama de casos de uso
- Diagrama de secuencia
- Diagrama de colaboración

- Diagrama de despliegue
- Diagrama de componentes
- Diseño de pantallas/interfaces

4.2.2. CASOS DE USO

Las tablas mostradas a continuación describen los procesos generales del sistema, desde la grabación de los comandos, pasando por el procesado de la voz en la PC, el reconocimiento del comando, etc:

Figura 4.1 Diagrama de casos de uso general



Fuente [Elaboración Propia]

A continuación mostramos los casos de uso para cada uno de los eventos del sistema los cuales servirán de base para el desarrollo de las interfaces que se muestran al final de este capítulo:

Tabla 4.1 Caso de uso #1

CASO DE USO #1		
<p>Caso de uso: Grabación de nuevos comandos y modificación de la base de datos</p> <p>Actores: Usuario, Técnico</p> <p>Tipo: Secundario</p>		
<p>Descripción: Para que el sistema sea lo mas eficientemente posible el usuario deberá grabar con su voz todos y cada uno de los comandos, si el usuario desea controlar algún evento en el equipo de televisión que no este contemplado en el sistema de control estándar, este deberá cambiar base de datos y también asignar una salida por el puerto USB.</p>		
Flujo	Eventos de actor	Eventos del sistema
<ol style="list-style-type: none"> 1. Grabación de nuevos comandos de voz. 2. Extracción de características de la señal de voz. 3. Almacenamiento en la base de datos. 	<ol style="list-style-type: none"> 1. Activar micrófono. 2. Pronunciar los nuevos comandos de voz. 3. Hacer corresponder los nuevos comandos con su correspondiente salida lógica en el puerto USB que servirá de interfaz entre software y hardware. 	<ol style="list-style-type: none"> 1. Extracción de características de la señal de voz. 2. Almacenamiento en la base de datos. 3. Correspondencia de cada comando a su respectiva salida lógica.

Fuente [Elaboración Propia]

Tabla 4.2 Caso de uso #2

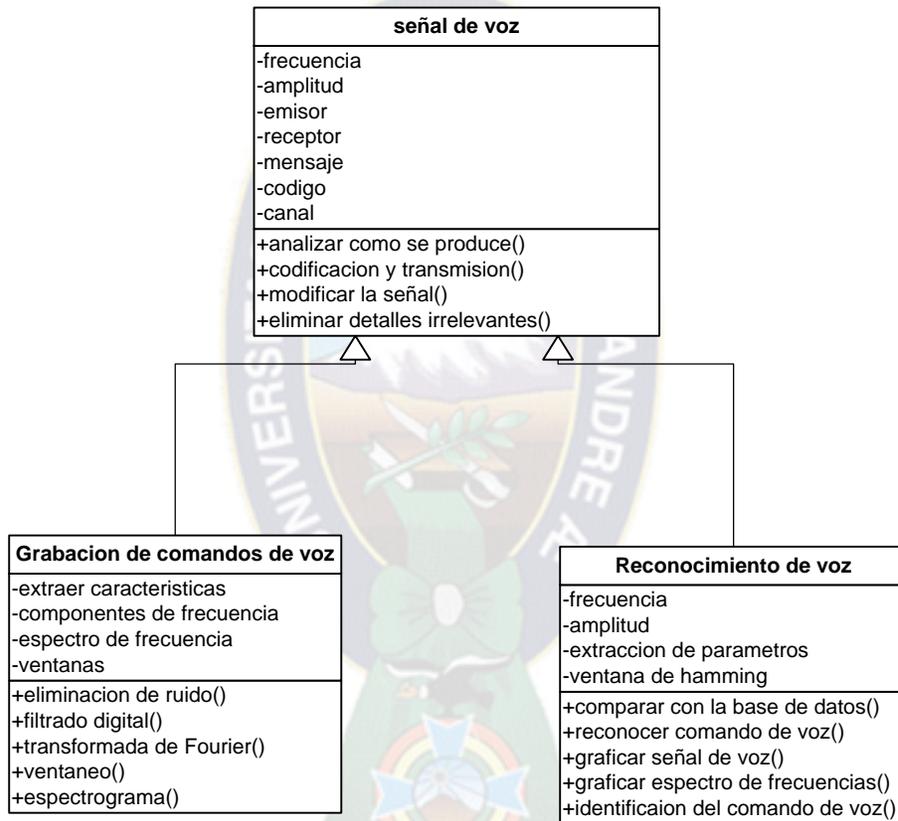
CASO DE USO #2		
Caso de uso: Reconocimiento		
Actores: Usuario		
Tipo: Primario		
Descripción: El usuario tiene la necesidad de activar algún evento del equipo de televisión, deberá pronunciar el comando de voz, si este es reconocido correctamente el sistema procede a activar el mismo.		
Flujo	Eventos de actor	Eventos del sistema
<ol style="list-style-type: none"> 1. Extracción de características de la voz 2. Medida de distancia 3. reconocimiento del comando 4. Activación del evento en el equipo de televisión ligado al comando de voz. 	<ol style="list-style-type: none"> 1. Elige que evento desea activar en el equipo de televisión. 2. Pronunciación del comando de voz correspondiente al evento deseado en el equipo de televisión. 	<ol style="list-style-type: none"> 1. Extracción de características del comando de voz. 2. Compara el comando de voz con la base de datos. 3. Reconoce el comando de voz emitido por el usuario. 4. Activa evento en el equipo de televisión.

Fuente [Elaboración Propia]

4.2.3. DIAGRAMA DE CLASES

La siguiente figura representa el diagrama de clases del sistema de control para un equipo de televisión basado en el reconocimiento de patrones de voz destinado a personas cuádruplégicas.

Figura 4.2 Diagrama de clases

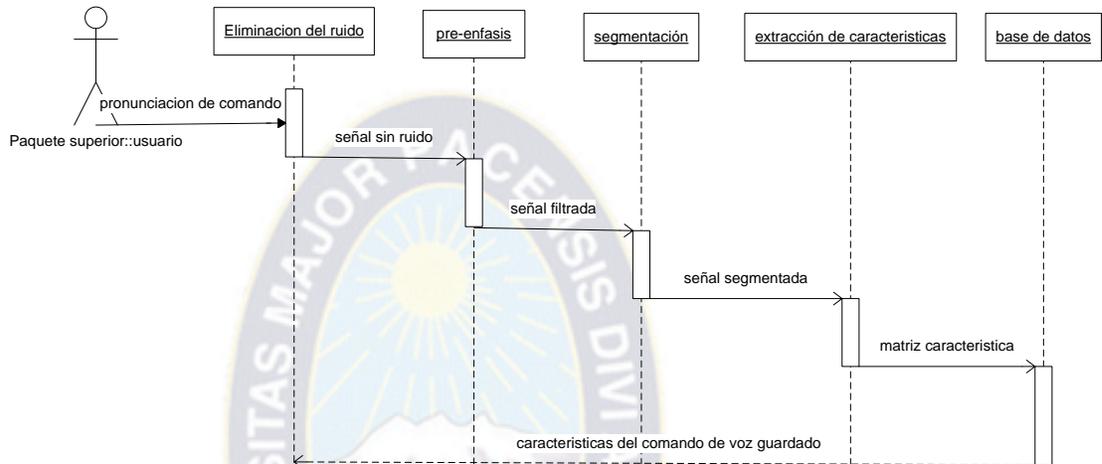


Fuente [Elaboración Propia]

4.2.4. DIAGRAMA DE SECUENCIA

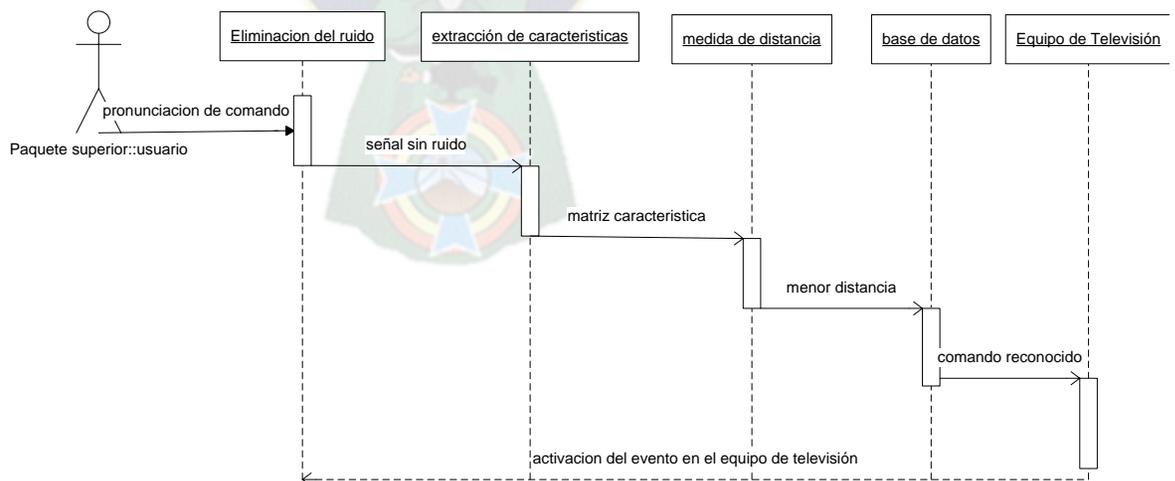
Los siguientes gráficos representan los procesos que sigue el sistema.

Figura 4.3 Grabación de comandos



Fuente [Elaboración Propia]

Figura 4.4 Reconocimiento de comando de voz

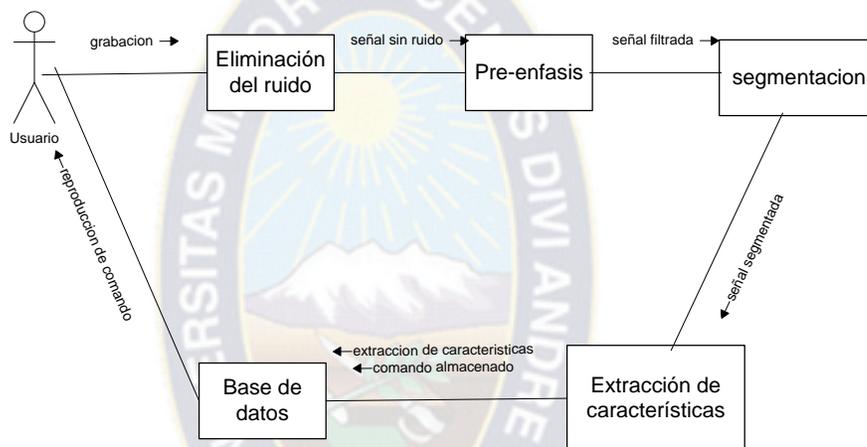


Fuente [Elaboración Propia]

4.2.5. DIAGRAMA DE COLABORACIÓN

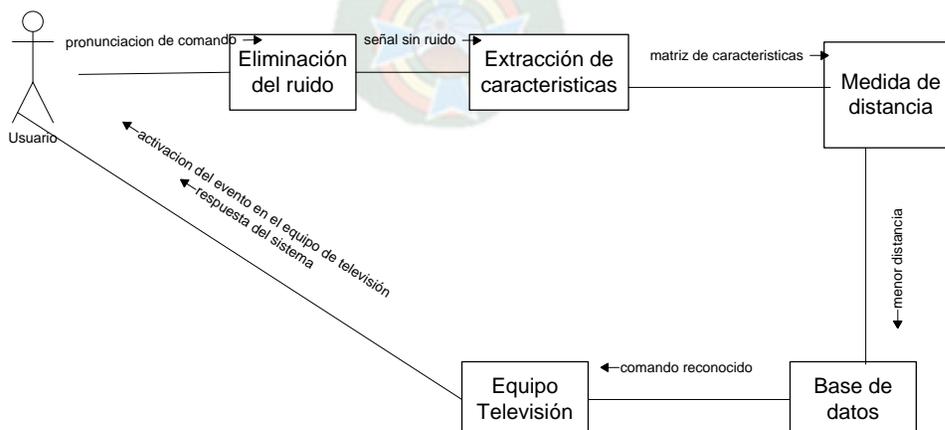
Este tipo de diagramas nos muestra una interacción organizada, basándose en los objetos que toman parte en la interacción y los enlaces entre los mismos (en cuanto a la interacción se refiere). A diferencia de los diagramas de secuencia, los diagramas de colaboración muestran las relaciones entre los roles de los objetos.

Figura 4.5 grabación de comandos



Fuente [Elaboración Propia]

Figura 4.6 Reconocimiento de comando de voz

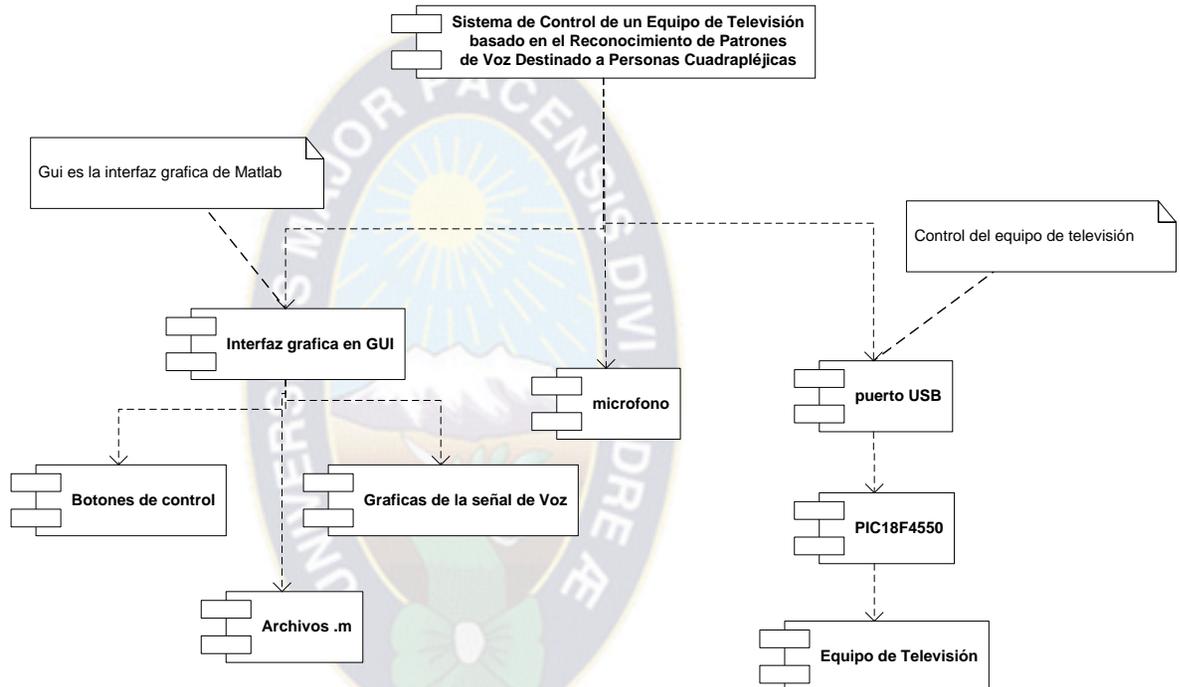


Fuente [Elaboración Propia]

4.2.6. DIAGRAMA DE COMPONENTES

El diagrama de componentes del software nos muestra la parte física del sistema, y se encuentra en la computadora estos son: la base de datos, archivos .m, interfaces graficas, interfaces hardware (USB), entre otros.

Figura 4.7 Diagrama de componentes

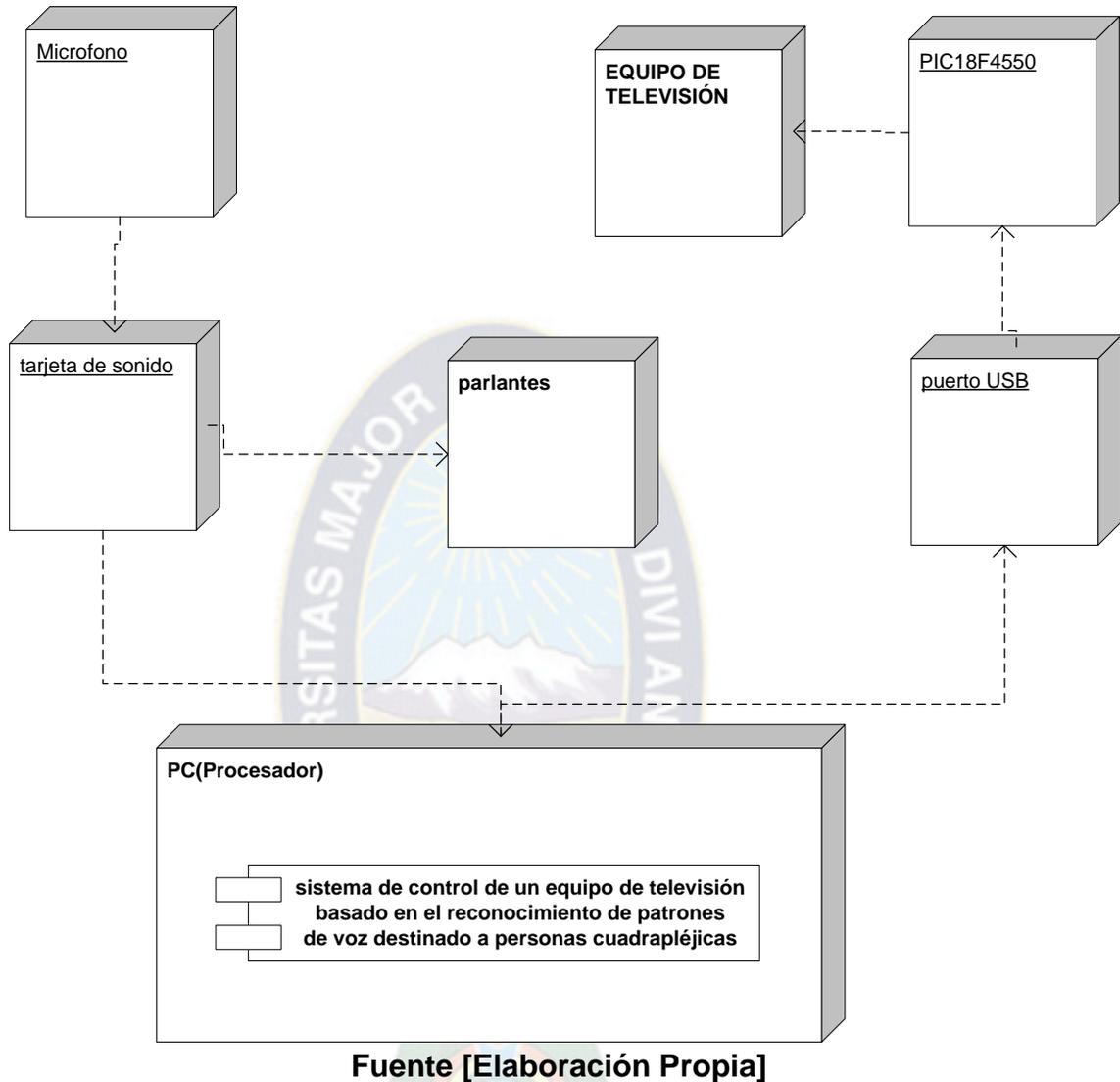


Fuente [Elaboración Propia]

4.2.7. DIAGRAMA DE DESPLIEGUE

Este tipo de diagramas se enfocan específicamente al hardware del sistema de control de un equipo de televisión basado en el reconocimiento de patrones de voz destinado a personas cuadrupléjicas.

Figura 4.8 Diagrama de despliegue



4.2.8. DISEÑO DE PANTALLAS/INTERFACES

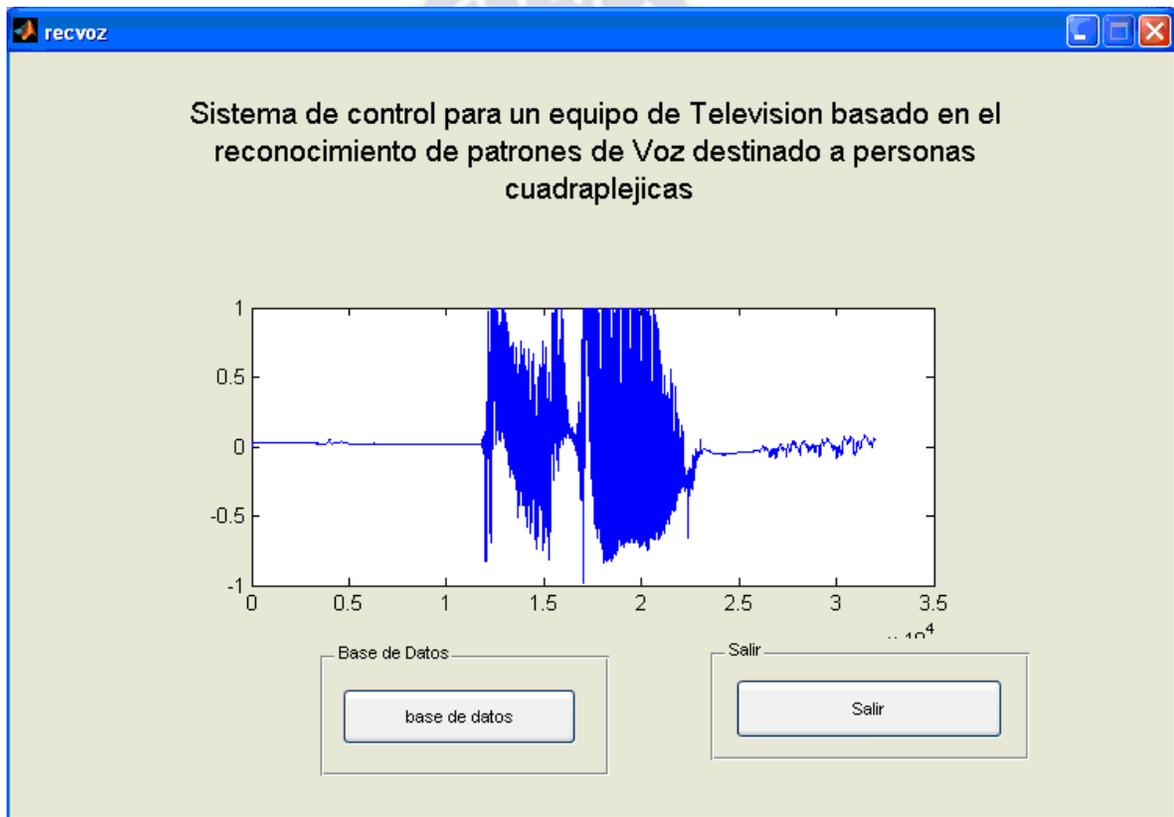
La interfaz final se realizó en Matlab 7.10 por su gran versatilidad en el manejo de vectores y matrices. Específicamente se utilizó una interfaz gráfica que es el Guide de Matlab.

GUIDE es un entorno de programación visual disponible en MATLAB para realizar y ejecutar programas que necesiten ingreso continuo de datos.

Tiene las características básicas de todos los programas visuales como Visual Basic o Visual C++.

La siguiente figura representa la pantalla principal del sistema en el se pueden observar la señal de voz en tiempo real de un comando por reconocer, también los enlaces hacia otras ventanas y la posibilidad de salir del sistema.

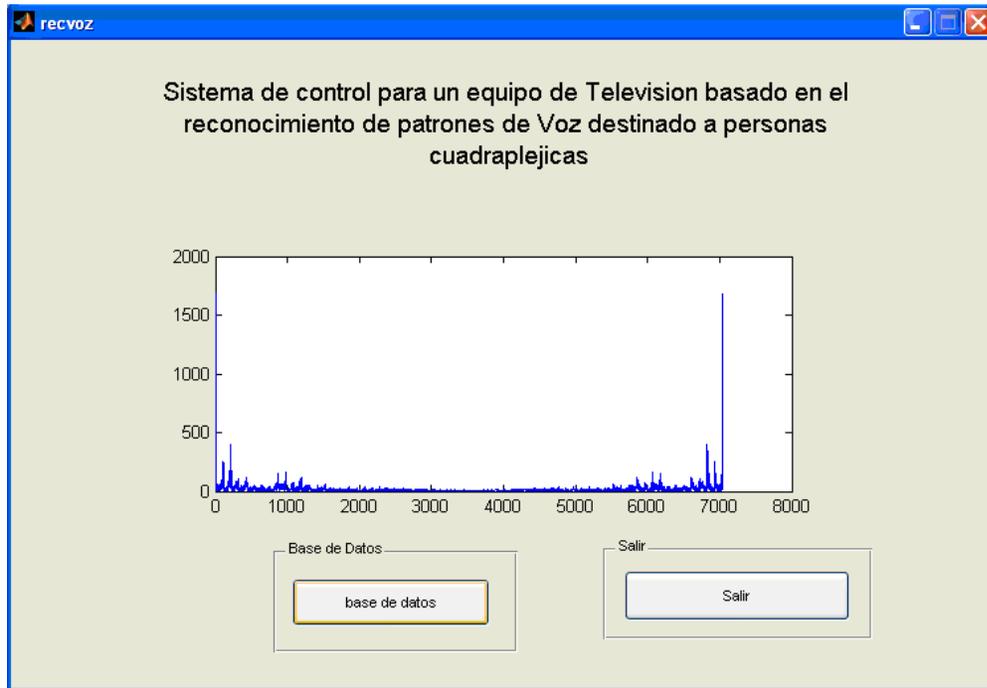
Figura 4.9 Pantalla principal grabación del comando



Fuente [Elaboración Propia]

La siguiente pantalla muestra el espectro de frecuencias de un comando ya reconocido.

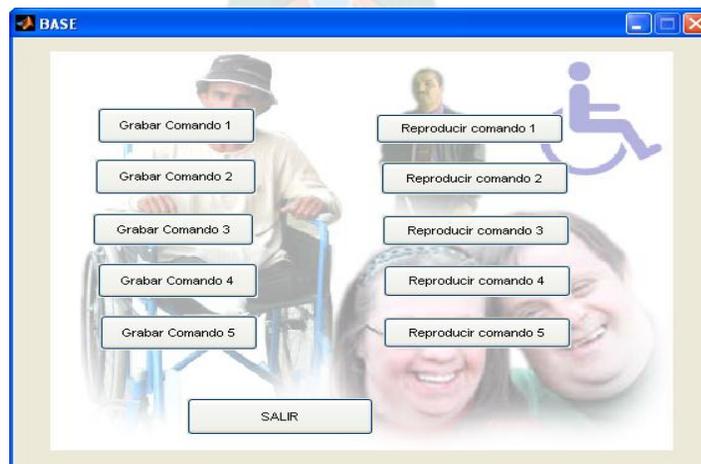
Figura 4.10 Pantalla principal mostrando el espectro de frecuencias



Fuente [Elaboración Propia]

La siguiente pantalla nos muestra la base de datos, en esta pantalla podemos modificar los comandos de voz y también reproducirlos.

Figura 4.11 Pantalla de base de datos



Fuente [Elaboración Propia]

4.3. COMUNICACIÓN USB

Las interfaces paralelo y serie RS-232 están desapareciendo prácticamente de los ordenadores personales y esto supone un problema, ya que muchas de las aplicaciones informáticas utilizan estos buses para la comunicación con el mundo exterior. La solución ideal es migrar a una interfaz USB y existen distintas formas de hacerlo. El método mas sencillo es emular RS-232 con el USB, con la ventaja de que el PC vera la conexión USB como una conexión COM RS-232 y no requerirá cambios en el software existente. Otra ventaja es que se utilizan drivers suministrados por Windows, por lo que no es necesario desarrollar uno propio; estos drivers son el usbser.sys y el cport.sys. Además, puesto que el protocolo USB maneja comunicaciones de bajo nivel, los conceptos baud rate, bit de paridad y control de flujo para le RS-232 ya no importan.

Por lo anterior expuesto se opto por la comunicación USB que será la interfaz entre software y hardware, esto se logra utilizando un microcontrolador que es el PIC18F4550 de MICROCHIP para mas características **[Ver Anexo D]**. Para la programación del microcontrolador utilizaremos el CCS que es un software para programar el mismo cuyo lenguaje es el C y para la simulación utilizaremos el ISIS de PROTEUS 7.

4.3.1. USB CDC(Communication Device Class)

Una clase USB es una agrupación de dispositivos de características comunes, es decir, utilizan una misma forma de comunicarse con el entorno. La clase de dispositivo permite conocer la forma en que la interfaz se comunica con el sistema, el cual puede localizar el driver que puede controlar la conectividad entre la interfaz y el sistema.

USB solo permite al driver comunicarse con el periférico a través de las tuberías (pipes) establecidas entre el sistema USB y los endpoints del periférico. Los tipos de transferencia a través de las pipes dependen del

endpoint y pueden ser: Bulk, Control, Interrupt e Isochronous. Una tubería es un enlace virtual entre el Host y el dispositivo USB, donde se configura el ancho de banda, el tipo de transferencia, la dirección del flujo de datos y el tamaño del paquete de datos.

Estos enlaces se definen y crean durante la inicialización del USB. Un endpoint es un buffer dentro del dispositivo o periférico donde se almacenan paquetes de información; todos los dispositivos deben admitir el endpoint 0, el cual recibe el control y las peticiones de estado durante la enumeración del dispositivo. Cuando se conecta un dispositivo al HOST se produce la enumeración en la cual el HOST interroga al dispositivo sobre sus características principales, asignándole una dirección y permitiendo la transferencia de datos.

La especificación clase de dispositivo de comunicación (CDC) define algunos modelos de comunicación, incluyendo la comunicación serie. Windows suministra el driver `subser.sys` para esta especificación. Para la especificación CDC se necesitan dos interfaces USB, primero la interfaz Communication Class usando un IN interrupt endpoint de interrupción y el segundo es la interfaz Data Class usando un OUT bulk endpoint y un IN bulk endpoint. Esta interfaz es utilizada para transferir los datos que normalmente deberían ser transferidos a través de la interfaz RS-232.

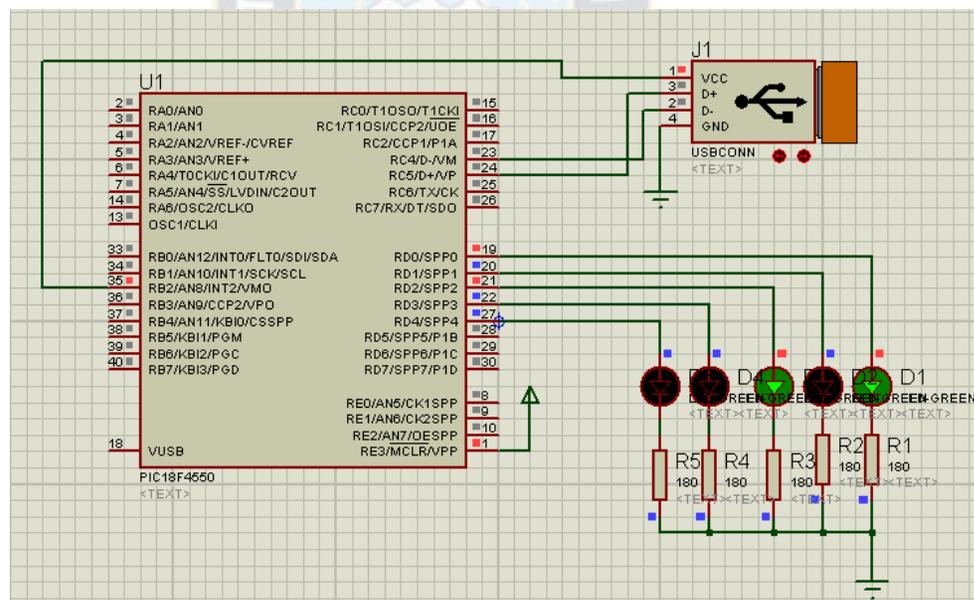
Desde el punto de vista del sistema USB, el dispositivo puede tener distintas configuraciones, para cada una de las cuales puede funcionar de forma distinta. Los dispositivos suministran la información necesaria al sistema USB a través de los descriptores; estos contienen unos campos que permiten al sistema clasificar al dispositivo y asignarle un driver. La primera información que necesita es la del fabricante y producto (USB vendor ID-VID y el Product ID-PID). El VID es un número de 16 bits por el USB Implementers Forum (USB-IF) y debe ser obtenido por el fabricante del dispositivo USB; cada VID

puede contener 65.536 PID diferentes al ser también un número de 16 bits. Microchip suministra su VIP y los PID para cada familia de PIC con USB.

Microsoft Windows XP no tiene un fichero *.inf estandar para el driver CDC, así que es necesario suministrar este fichero cuando se conecta un dispositivo USB por primera vez al sistema. Microchip suministra el fichero mchpcdc.inf necesario para sus dispositivos PIC.

El siguiente circuito es la simulación de la comunicación USB entre la PC y el mundo exterior específicamente el control remoto del equipo de televisión, los leds a la salida del puerto B representan a los eventos en el equipo de televisión.

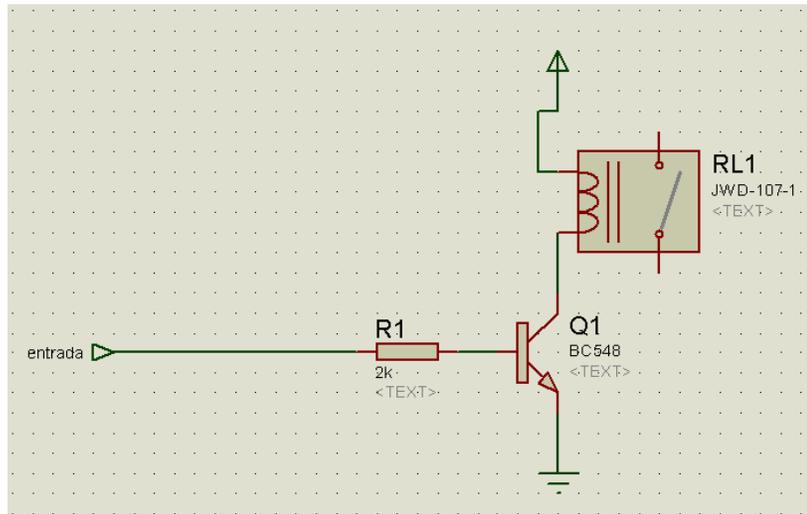
Figura 4.12 Circuito simulado en ISIS



Fuente [Elaboración propia]

La siguiente figura representa el circuito de control, la entrada de este es una de las salidas del microcontrolador PIC18F4550, el transistor trabaja en corte y saturación haciendo que el rele se active o desactive, el cual a su vez activa un evento deseado en el equipo de televisión.

Figura 4.13 Circuito simulado en ISIS



Fuente [Elaboración propia]

4.3.2. PROGRAMA DEL MICROCONTROLADOR PIC18F4550

Seguidamente se muestra el software implementado en el microcontrolador PIC18F4550 el cual fue elaborado en PIC C Compiler que es un lenguaje para microcontroladores basado en C.

```
#include <18F4550.h>
#fuses
HSPLL,NOWDT,NOPROTECT,NOLVP,NODEBUG,USBDIV,PLL5,CPUDIV1,VRE
GEN
#use delay(clock=48000000)
#define USB_CON_SENSE_PIN PIN_B2
#include <usb_cdc.h>
#define LED1 PIN_D0
#define LED2 PIN_D1
#define LED3 PIN_D2
#define LED4 PIN_D3
#define LED5 PIN_D4
#define LED_ON output_high
#define LED_OFF output_low
void main()
{
char rx;
int cont5=0;
setup_adc_ports(NO_ANALOGS);
```

```

LED_OFF(LED1);
LED_OFF(LED2);
LED_OFF(LED3);
LED_OFF(LED4);
LED_OFF(LED5);

usb_init();
usb_cdc_init();

while(!usb_cdc_connected());

do
{
    usb_task();
    if(usb_enumerated())
    {
        delay_ms(300);
        rx=usb_cdc_getc();
        switch (rx)
        {
            case '1':
                LED_ON(LED1);
                delay_ms(300);
                LED_OFF(LED1);
                break;

            case '2':
                LED_ON(LED2);
                delay_ms(300);
                LED_OFF(LED2);
                break;

            case '3':
                LED_ON(LED3);
                delay_ms(300);
                LED_OFF(LED3);
                break;

            case '4':
                LED_ON(LED4);
                delay_ms(300);
                LED_OFF(LED4);
                break;

            case '5':
                if(cont5%2==0)

```



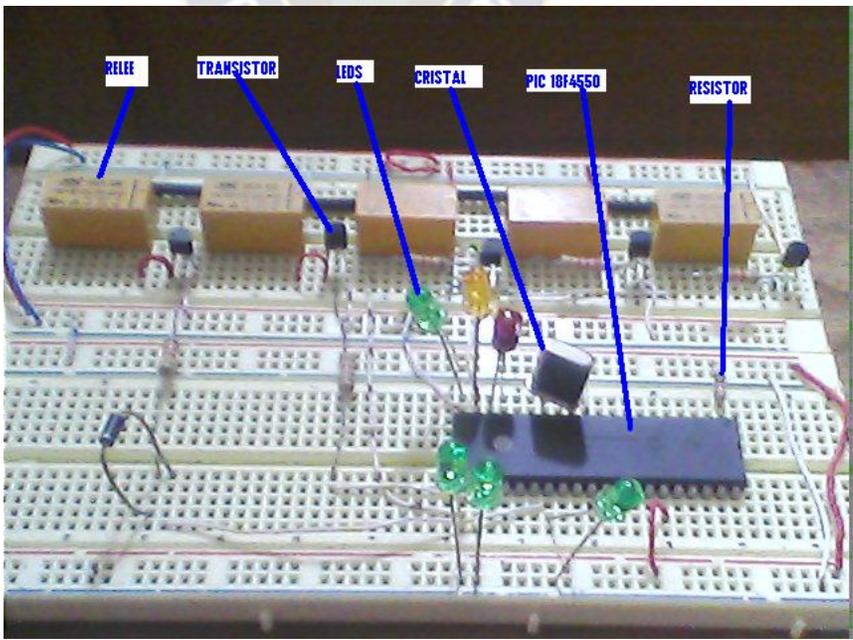
```
{
  LED_ON(LED5);
  cont5++;
}
else
{
  LED_OFF(LED5);
  cont5++;
}

break;

default:break;
}
}
while(true);
}
```

Finalmente en la siguiente figura se muestra el circuito armado en un protoboard.

Figura 4.14 Circuito real armado en un protoboard



Fuente [Elaboración propia]

4.4. PRUEBAS Y RESULTADOS

A continuación se muestran los resultados obtenidos del proyecto, por ser el primer prototipo se analizan cinco comandos de voz pronunciados por cinco diferentes personas, se analizan los resultados obtenidos para obtener el factor de calidad del sistema. Se realizan pruebas de caja negra, para dichas se selecciono a usuarios finales ya que son ellos los que conocen mas de los requerimientos, las restricciones y los resultados que debería proporcionar el sistema.

4.4.1. PRUEBAS DEL SOFTWARE

En la siguiente tabla se muestra la respuesta del software a cinco comandos pronunciados por cinco diferentes personas. Los comandos están en ingles debido a que en este idioma las palabras son más cortas en su pronunciación y son más fáciles de reconocer para el sistema, estos comandos son: UP, DOWN, NEXT, PREV y TV que corresponden a los siguientes eventos: aumentar volumen, disminuir volumen, siguiente canal, anterior canal y encendido ó apagado del televisor.

Tabla 5.1 Pruebas de reconocimiento de voz

Nro. de persona	Comando pronunciado	Distancia	Comando reconocido	Observación
1	UP	47.8032	UP	Comando reconocido correctamente
		148.832		
		162.404		
		138.522		
		232.238		
	DOWN	135.019	DOWN	Comando reconocido correctamente
		57.5854		
		158.213		

		67.989		
		179.103		
	NEXT	168.885	NEXT	Comando reconocido correctamente
		156.376		
		86.8046		
		144.865		
		149.517		
	PREV	87.1282	PREV	Comando reconocido correctamente
		81.2845		
		140.178		
		66.2354		
		213.538		
	TV	238.264	TV	Comando reconocido correctamente
		194.798		
		161.295		
		360.121		
		52.0473		
2	UP	85.2383	UP	Comando reconocido correctamente
		135.184		
		166.836		
		131.648		
		263.165		
	DOWN	117.683	DOWN	Comando reconocido correctamente
		63.7105		
		147.915		
		69.6386		
		182.455		
	NEXT	224.592	NEXT	Comando reconocido correctamente
		225.184		
		186.07		

		190.43		
		196.897		
	PREV	74.6608	PREV	Comando reconocido correctamente
		86.301		
		147.424		
		64.884		
		175.178		
	TV	208.316	TV	Comando reconocido correctamente
		191.752		
		123.452		
		327.08		
		59.1468		
3	UP	76.1999	UP	Comando reconocido correctamente
		118.315		
		152.409		
		121.6		
		139.2		
	DOWN	124.218	DOWN	Comando reconocido correctamente
		61.0778		
		113.83		
		73.8354		
		156.195		
	NEXT	171.536	NEXT	Comando reconocido correctamente
		189.845		
		85.2689		
		169.605		
		157.592		
	PREV	93.9249	UP	Comando reconocido incorrectamente
		148.104		
		186.629		

		120.917		
		226.751		
	TV	230.235	TV	Comando reconocido correctamente
		186.586		
		152.809		
		370.04		
		53.5074		
4	UP	61.8422	UP	Comando reconocido correctamente
		146.03		
		183.509		
		143.786		
		257.53		
	DOWN	147.702	DOWN	Comando reconocido correctamente
		55.9169		
		150.876		
		73.3332		
		169.782		
	NEXT	153.905	NEXT	Comando reconocido correctamente
		201.71		
		73.9852		
		173.453		
		163.484		
	PREV	89.6462	PREV	Comando reconocido correctamente
		111.519		
		173.528		
		86.6424		
		200.275		
TV	232.629	TV	Comando reconocido correctamente	
	182.513			
	157.288			

		347.702		
		52.2128		
5	UP	58.2887	UP	Comando reconocido correctamente
		167.224		
		193.746		
		146.888		
		245.994		
	DOWN	126.947	DOWN	Comando reconocido correctamente
		54.7178		
		179.344		
		70.1288		
		203.268		
	NEXT	136.32	NEXT	Comando reconocido correctamente
		110.924		
		66.8229		
		104.506		
		130.556		
	PREV	93.405	PREV	Comando reconocido correctamente
		117.826		
		179.228		
		88.2356		
		218.264		
TV	267.316	NEXT	Comando reconocido incorrectamente	
	357.715			
	117.329			
	347.767			
	149.677			

Fuente [Pruebas del sistema]

Del número total de eventos solo dos son reconocidos erróneamente por lo que tenemos los siguientes resultados:

$$\frac{23}{25} * 100\% = 92\%$$

Se tiene un factor de calidad de 92%, por lo que el sistema es aceptable y cumple con los objetivos propuestos en el capítulo I.

4.5. MÉTRICAS MC. CALL

Para las métricas de Mc. Call se procedió al llenado de un cuestionario con diferentes preguntas [Ver Anexo C].

A continuación se muestran los resultados a los cuestionarios.

Tabla 5.2 Resultados de los cuestionarios

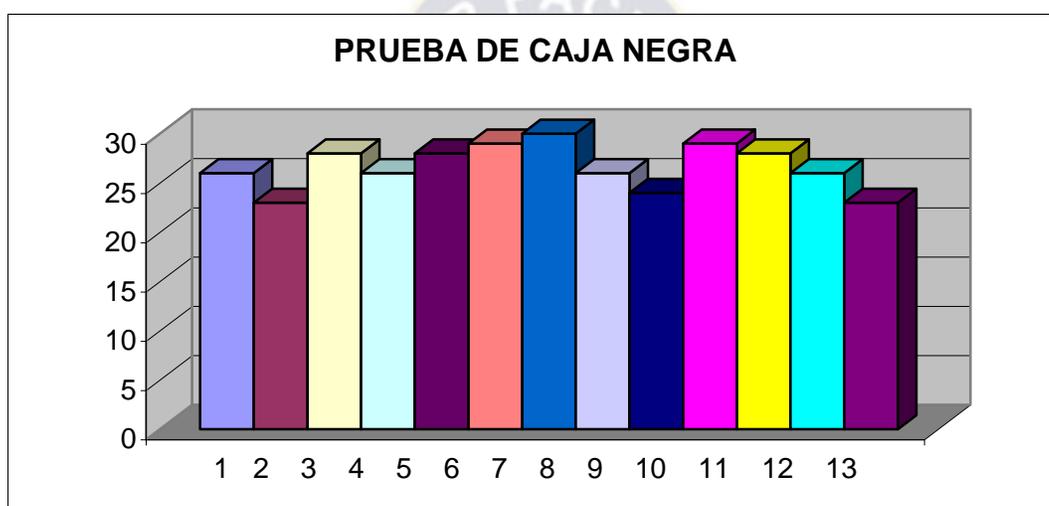
Nro.	Métricas de Mc. Call	Evaluación						Resultado Final	
		0	1	2	3	4	5		6
1	Complejión					1	2	2	26
2	Consistencia					2	3		23
3	Estandarización de datos						2	3	28
4	Tolerancia al error						4	1	26
5	Eficiencia de ejecución						2	3	28
6	Generalidad						1	4	29
7	Exactitud							5	30
8	Estandarización de comunicaciones						4	1	26
9	Instrumentación					2	2	1	24
10	Modularidad						1	4	29
11	Operatividad						2	3	28
12	Seguridad					1	2	2	26

13	Formación				2	3	23
Resultado obtenido/30puntos							26.62
Valor porcentual de las pruebas sobre 100%							88.72

Fuente [cuestionarios]

El siguiente grafico muestra en resumen los resultados obtenidos del cuestionario.

Figura 5.1 prueba de caja negra



Fuente [Elaboración propia]

Basándonos en los resultados finales obtenidos anteriormente y promediando los mismos llegamos a la siguiente conclusión:

$$\frac{92 + 88.72}{2} = 90.36\%$$

En promedio general el factor de calidad es de 90.36%, situándose dentro de un rango aceptable. Por lo cual el sistema pasa las pruebas de caja negra, lo cual indica que el sistema es eficiente y confiable.



CAPÍTULO V

ANÁLISIS DE COSTOS

CAPÍTULO 5. ANÁLISIS DE COSTOS

5.1. INTRODUCCIÓN

El presente capítulo hace referencia a todos los gastos que se realizaron tanto en el desarrollo del sistema como en su implementación, se especificara detalladamente los siguientes puntos: software, hardware.

En el diseño e implementación del proyecto, en todo momento se tuvo en cuenta la economía tratando de utilizar herramientas y componentes que no tuvieran un gran coste pero cuidando que los mismos sean lo más eficientemente posible.

5.2. FACTIBILIDAD TÉCNICA

En este punto evaluaremos si tanto el equipo como el software están disponibles, y tienen las capacidades técnicas requeridas para cada alternativa del diseño que se este considerando.

5.2.1. COSTOS DE DESARROLLO

Para el desarrollo del sistema se empleo software y hardware que a continuación mencionaremos:

5.2.1.1. SOFTWARE

En este punto se menciona los programas que se utilizo en el desarrollo del proyecto, describiéndolos en sus versiones respectivas.

Cabe recalcar que el presente proyecto fue realizado en las instalaciones de la carrera de Electrónica y Telecomunicaciones específicamente en los laboratorios de la misma. En consecuencia los costos de software y hardware son nulos ya que la carrera cuenta con los equipos y licencias respectivas para su uso.

Tabla 6.1 Software para el desarrollo

Software
Sistema operativo Windows XP Spanish
Microsoft Office 2007
Matlab 7.10
PIC C Compiler
Proteus 7.6

Fuente [propia]

5.2.1.2. HARDWARE

A continuación detallaremos los equipos y componentes que se utilizaron para llevar a cabo el presente proyecto. Cabe recalcar que algunos equipos no tienen costo ya que la carrera de Electrónica y Telecomunicaciones cuenta con los mismos.

Tabla 6.2 Hardware para el desarrollo

Hardware	Cantidad	Precio por unidad(\$us)	Precio
<p><i>PC</i></p> <ul style="list-style-type: none"> • Microprocesador Intel Pentium D 3 Ghz • Placa madre D955XC • Memória DDR2 SDRAM de 512 MB • Tarjeta de vídeo PCI Express • Tarjeta de sonido High definition subsystem 	1	-	-

<ul style="list-style-type: none"> • Disco duro 160 Gb • Monitor de 19 pulgadas • Case Delux 			
Microcontrolador PIC 18F4550	1	15	15
Cable USB	1	2	2
Micrófono Omega	1	8	8
Grabador de PICs	1	-	-
Oscilador Cristal 20Mhz	1	2	2
Reles de 5V	8	1	8
Fuente de alimenticio regulable	1	-	-
Total(\$us)			35
Total(Bs) T.C.=7.07			247.45

Fuente [Cotizaciones]

5.3. FACTIBILIDAD OPERATIVA

La facilidad de uso que tendrá el prototipo es de solo necesitar de la pronunciación de un comando en específico por parte del usuario y el monitoreo del mismo por parte del sistema para activar un evento en el equipo de televisión.

Su fácil operabilidad se da ya que el sistema analiza el comando pronunciado y compara con la base de datos existente lo cual da como resultado el evento esperado en el equipo de televisión.

5.4. COSTO DEL SOFTWARE DE RECONOCIMIENTO DE VOZ

El costo del software lo definiremos basándonos en una métrica de costos que es COCOMO intermedio ya que proporciona una manera potente de capturar la influencia del entorno en el proyecto.

En este modelo se introducen 15 atributos de coste para tener en cuenta el entorno de trabajo. Estos atributos se utilizan para ajustar el coste nominal del proyecto al entorno real, incrementando la precisión de la estimación.

5.4.1. ECUACIONES NOMINALES DE COSTE

Para cada modo de desarrollo, los 15 atributos del coste intervienen como multiplicadores en el coste nominal, $M(x)$, para producir el coste ajustado.

Las ecuaciones de COCOMO en este modo Intermedio son:

$$\begin{aligned} E &= a \cdot KLDC^b \cdot M_{(x)} \\ T &= c \cdot E^d \\ P &= \frac{E}{T} \end{aligned} \quad \text{Ec. 5.1}$$

Donde E es el esfuerzo aplicado en persona-mes, T es el tiempo de desarrollo en meses, $KLDC$ es el número de líneas estimadas para el proyecto (en miles) y P es el número de personas necesarias. Los coeficientes a , b , c y d se obtienen de la siguiente tabla:

Tabla 6.3 Coeficientes COCOMO Intermedio

Modo	a	b	c	d
Orgánico	3.2	1.05	2.5	0.38
Semiacoplado	3.0	1.12	2.5	0.35
Empotrado	2.8	1.2	2.5	0.32

Fuente [<http://www.softwaremetrics.com>]

Para la estimación de costos en el presente proyecto utilizaremos el modo empotrado ya que es un proyecto software que se debe desarrollar con requisitos hardware, software y de operación.

5.4.2. ATRIBUTOS DE COSTE

Estos atributos tratan de capturar el impacto del entorno del proyecto en el coste de desarrollo. De un análisis estadístico de más de 100 factores que influyen el coste, Boehm retuvo 15 de ellos para COCOMO.

Estos atributos se agrupan en cuatro categorías: atributos del producto, atributos del personal y atributos del proyecto.

5.4.2.1. ATRIBUTOS DEL PRODUCTO

RELY: garantía de funcionamiento requerida al software

DATA: tamaño de la base de datos

CPLX: complejidad del producto

5.4.2.2. ATRIBUTOS DEL PERSONAL

ACAP: capacidad del analista

AEXP: experiencia en la aplicación

PCAP: capacidad del programador

VEXP: experiencia en máquina virtual

LEXP: experiencia en lenguaje de programación

5.4.2.3. ATRIBUTOS DEL PROYECTO

MODP: prácticas de programación modernas

TOOL: utilización de herramientas software

SCED: plan de desarrollo requerido.

Cada atributo se cuantifica para un entorno de proyecto. La escala es:

- muy bajo
- bajo
- nominal
- alto
- muy alto
- extremadamente alto

En la tabla se muestran los valores del multiplicador para cada uno de los 15 atributos. Estos 15 valores se multiplican.

Tabla 6.4 Valores de los atributos

Grado	Muy Baja	Baja	Nominal	Alta	Muy Alta	Extra Alta
Atributos del Producto						
RELY	0.75	0.88	1.00	1.15	1.40	
DATA		0.94	1.00	1.08	1.16	
CPLX	0.70	0.85	1.00	1.15	1.30	1.65
Atributos del Hardware						
TIME			1.00	1.11	1.30	1.62
STOR			1.00	1.06	1.21	1.56
VIRT		0.87	1.00	1.15	1.30	
TURN		0.87	1.00	1.07	1.15	
Atributos del Personal						
ACAP	1.46	1.19	1.00	0.86	0.71	
AEXP	1.29	1.13	1.00	0.91	0.82	
PCAP	1.42	1.17	1.00	0.86	0.70	
VEXP	1.21	1.10	1.00	0.90		
LEXP	1.14	1.07	1.00	0.95		
Atributos del Proyecto						
MODP	1.24	1.10	1.00	0.91	0.82	
TOOL	1.24	1.10	1.00	0.91	0.83	
SCED	1.23	1.08	1.00	1.04	1.10	

Fuente [<http://www.softwaremetrics.com>]

5.4.3. SIGNIFICADO DE LOS ATRIBUTOS

RELY

Indica las posibles consecuencias para el usuario en el caso que todavía existan defectos en el producto. Una puntuación 'muy baja' indica que solamente hace falta eliminar los defectos sin ninguna otra consecuencia.

"muy bajo": el efecto de un fallo del software simplemente trae como consecuencia la inconveniencia de corregir el fallo

"bajo": el efecto de un fallo software es una pérdida fácilmente recuperable para los usuarios

"nominal": el efecto es una moderada pérdida para los usuarios, pero es una situación de la que se puede salir sin excesiva dificultad

"alto": el efecto es una gran pérdida financiera o una inconveniencia masiva humana

"muy alto": el efecto es una pérdida de vidas humanas.

Para el presente proyecto el efecto de un fallo del software simplemente trae como consecuencia la inconveniencia de corregir el fallo por lo tanto se toma como muy bajo entonces $RELY=0.75$

DATA

Indica el tamaño de la base de datos a desarrollar en relación con el tamaño del programa. Tenemos cuatro segmentos con la razón 10-100-1000, que determinan las puntuaciones de "bajo" a "muy alto".

En el presente proyecto tomamos $DATA=0.94$

CPLX

Indica la complejidad de cada módulo y se utiliza para determinar la complejidad compuesta del sistema. Entonces la puntuación puede variar de "muy bajo" a "extra alta".

Debido a que en el presente proyecto se utilizan archivos .m que a su vez invocan a otros archivos .m la complejidad es muy alta por lo que tomamos $CPLX=1.30$

TIME

Siempre será más exigente para un programador escribir un programa que tiene una restricción en el tiempo de ejecución. Esta puntuación se expresa en el porcentaje de tiempo de ejecución disponible. Es 'nominal' cuando el

porcentaje es el 50%, y 'extremadamente alto' cuando la restricción es del 95%.

En el presente proyecto la respuesta del sistema debe ser lo mas rápida posible por lo tanto tomamos $TIME=1.30$.

STOR

Se espera que un cierto porcentaje del almacenamiento principal sea utilizado por el programa. El esfuerzo de programación se incrementa si el programa tiene que correr en un volumen menor del almacenamiento principal. STOR captura este esfuerzo extra de 'nominal' cuando la reducción del almacenamiento principal es del 50% a 'extremadamente alto' cuando la reducción es del 95%.

Para el presente proyecto se toma $STOR=1$ ya que debido al avance de la tecnología los volúmenes de memoria RAM y ROM son muy grandes y no es muy necesario reducir o comprimir el software.

VIRT

Durante el desarrollo del software la máquina (hard y soft) en la que el programa se va a desarrollar puede sufrir algunos cambios. VIRT lo refleja desde 'bajo' a 'muy alto'.

En el presente proyecto la maquina en la que se desarrolla el sistema sufre cambios de software y hardware ya que adhieren elementos hardware y se tuvieron que instalar software y drivers por lo que $VIRT=1.15$

TURN

Cuantifica el tiempo de respuesta del ordenador desde el punto de vista del programador. Cuanto mayor sea el tiempo de respuesta, más alto será el esfuerzo humano. TURN puede variar desde 'bajo' para un sistema interactivo a 'muy alto', cuando el tiempo medio de respuesta es de más de 12 horas.

El presente proyecto es un proyecto interactivo por lo tanto tomamos $TURN=0.87$.

ACAP

La capacidad del grupo de analistas, en términos de habilidad de análisis, eficiencia y capacidad para cooperar tiene un impacto significativo en el esfuerzo humano. Cuanto más capaz sea el grupo, menos esfuerzo será necesario. ACAP puede variar desde 'muy bajo' a 'muy alto'.

En el presente proyecto se toma $ACAP=0.86$.

AEXP

La experiencia del grupo en una aplicación similar tiene una gran influencia en el esfuerzo. Puede variar desde 'muy bajo' (menos de cuatro meses de experiencia) a 'muy alto' (mayor de 12 años de experiencia).

"muy bajo": < 4 meses experiencia media

"bajo": 1 año de experiencia media

"nominal": 3 años de experiencia media

"alto": 6 años de experiencia media

"muy alto": > 12 años, o reimplementación de un subsistema

En el presente proyecto se toma $AEXP=1.13$ ya que se tuvieron experiencias anteriores mas o menos alrededor de 1 año.

PCAP

La cuantificación es similar a la de ACAP, pero en este caso relacionado con los programadores. Se aplica a los programadores como grupo, pero no a los programadores individuales.

Para el presente proyecto se toma $PCAP=0.86$

VEXP

Cuanto mayor sea la experiencia del grupo de programación con el

procesador, menor será el esfuerzo necesario. VEXP puede variar desde 'muy bajo', cuando la experiencia es menor de un mes, a 'alto' cuando esta experiencia es mayor de 3 años.

"muy bajo": < 1 mes experiencia media

"bajo": 4 meses

"nominal": 1 año

"alto": > 3 años

Debido a que anteriormente a la implementación del proyecto se tuvieron experiencias alrededor de 1 año con proyectos similares se toma $VEXP=1$.

LEXP

Un grupo de programadores con amplia experiencia en un lenguaje determinado programará de una manera mucho más segura, generando un menor número de defectos y de requerimientos humanos. Puede variar desde 'muy bajo' a 'alto' para un grupo de un mes a tres años de experiencia, respectivamente.

"muy bajo": < 1 mes experiencia media

"bajo": 4 meses de experiencia media

"nominal": 1 año de experiencia media

"alto": > 3 años

Se toma $LEXP=0.95$ debido a que se tiene experiencia con Matlab que es el lenguaje fundamental para el presente proyecto.

MODP

Utilización de modernas prácticas de programación. Varía de 'muy bajo' a 'muy alto'. Estas prácticas incluyen, por ejemplo, programación estructurada y desarrollo 'top-down'.

"muy bajo": no se utilizan prácticas modernas de programación -PMP-.

"bajo": uso experimental de algunas PMP

"nominal": experiencia razonable en el uso de algunas PMP

"alto": experiencia razonable en gran parte de PMP

"muy alto": uso habitual de PMP

Tomamos MODP=0.91 ya que los algoritmos en Matlab utilizan modernas técnicas de programación.

TOOL

El uso adecuado de herramientas software es un multiplicador de la productividad. La puntuación de TOOL varía desde 'muy bajo' cuando sólo se utilizan herramientas básicas, a 'muy alto' cuando se utilizan herramientas específicas.

En el presente proyecto se utilizan herramientas específicas y de alta complejidad por lo que TOOL=0.83.

SCED

El tiempo nominal de desarrollo, tal como se define en el modo básico, es el plazo que requiere menor esfuerzo humano. Cualquier apresuramiento ('muy bajo') o retraso ('muy alto') demandarán más esfuerzo.

En el presente proyecto se tuvieron que apresurar algunos puntos en el diseño e implementación por lo que SCED=1.08.

5.4.4. CÁLCULO DEL ESFUERZO, TIEMPO Y NÚMERO DE PERSONAS

Con todos los atributos ya determinados en el anterior paso continuamos con el cálculo del valor de M(X)

$$M_{(x)} = 0.75 * 0.94 * 1.30 * 1.30 * 1 * 1.15 * 0.87 * 0.86 * 1.13 * 0.86 * 1 * 0.95 * 0.91 * 0.83 * 1.08$$

$$M_{(x)} = 0.772$$

Las líneas de código de los archivos .m del presente proyecto suman en total 874, con este dato proseguimos con los cálculos:

$$\text{Si } LDC=874 \Rightarrow KLDC = \frac{LDC}{1000} = \frac{874}{1000} = 0.874$$

Cálculo del Esfuerzo

$$E = a \cdot KLDC^b \cdot M_{(x)} = 2.8 \cdot (0.874)^{1.2} \cdot 0.772 = 1.84 \approx 2 \text{ [personas - mes]}$$

Cálculo del Tiempo

$$T = c \cdot E^d = 2.5 \cdot 2^{0.32} = 3.12 \approx 3 \text{ [meses]}$$

Cálculo del número de personas

$$P = \frac{E}{T} = \frac{2}{3} = 0.667 \approx 1 \text{ [persona]}$$

5.4.5. ESTIMACIÓN DEL COSTO DEL SOFTWARE

Para obtener el costo del software fue necesario saber cuanto gana un programador por mes. Para esto se hizo una pequeña encuesta en empresas que se dedican al desarrollo de software en la ciudad de La Paz.

Tabla 6.5 Salario de un programador por mes

Empresa	Sueldo-mes(\$us)
Quiroga System	400
Neurosystem	350
Creaxium	450
Promedio	SP=400

Fuente [Encuesta a profesionales de la rama]

El salario promedio de un programador en nuestro medio es de \$us400 por mes, Prosiguiendo con nuestros cálculos tenemos que:

Cálculo del costo del software(CS)

$$CS = P \cdot T \cdot SP = 1 \cdot 3 \cdot 400 = 1200$$

Por lo tanto el costo del software es de **\$us 1200**.

5.5. COSTO TOTAL DEL PROYECTO(CTS)

El costo total del proyecto lo calcularemos sumando costos software, costos hardware y el costo del desarrollo del software.

$$CTS = CS + Hw + Sw$$

Donde:

CS: costo de desarrollo de software

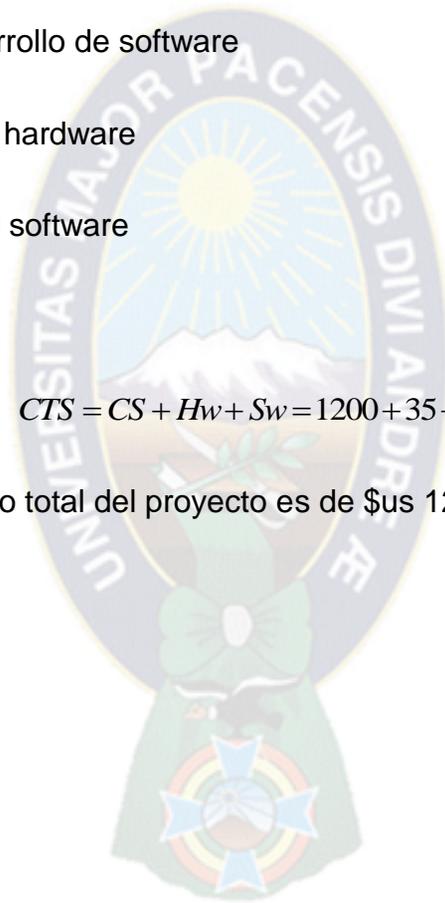
Hw: costo total del hardware

Sw: Costo total del software

Entonces:

$$CTS = CS + Hw + Sw = 1200 + 35 + 0 = 1235$$

Por lo tanto el costo total del proyecto es de \$us 1235.





CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES

6.1. CONCLUSIONES

Al culminar el presente proyecto de grado se obtuvieron las siguientes conclusiones:

- Concluimos que al implementar el sistema de control para un equipo de televisión se mejora la calidad de vida de los discapacitados cuadraplégicos. Ya que ellos ya no tienen la dificultad de operar su equipo de televisión.
- Con el presente proyecto se brinda la suficiente autonomía a las personas con discapacidad cuadraplégica, debido a que ellas ya pueden operar su equipo de televisión por sí mismas y no dependen de otras personas para hacerlo.
- Se mejora la comodidad de las personas con discapacidad cuadraplégica, proporcionándoles el sistema de control para su equipo de televisión por comandos de voz, logrando así que la vida de estas personas sea lo más placentera posible.
- El reconocimiento del habla representa una de las aplicaciones más importantes dentro del reconocimiento de patrones y el cual es posible llevar a cabo mediante diversas técnicas de inteligencia artificial. El considerar una técnica en especial, depende de las necesidades que pretende cada reconocedor, de forma que para el presente trabajo se requirió de un procesamiento rápido y eficaz tal como el que nos proporciono los coeficientes de predicción lineal.
- Los módulos diseñados e implementados en el presente trabajo desempeñan actividades de gran índole y mediante las cuales se logra mejorar el rendimiento dentro del sistema de reconocimiento de comandos de voz para el control del equipo de televisión.
- Matlab presenta una gran facilidad en el manejo de vectores y matrices es por esta razón que el software de reconocimiento de voz fue implementado en Matlab 7.10, logrando así obtener un software amigable para el usuario final.

- El resultado final del reconocimiento de comandos de voz es de un 90.36% de efectividad, por lo que el sistema es eficiente y confiable. Durante el desarrollo se comprueba que el sistema de reconocimiento de comandos de voz es rápida y eficaz.

6.2. RECOMENDACIONES

Se recomienda tener en cuenta los siguientes puntos al momento de ejecutar el sistema:

- Para que el sistema sea lo mas efectivo posible se recomienda que cada usuario grabe con su voz todos los comandos.
- Para futuro se recomienda incorporar un micrófono inalámbrico.

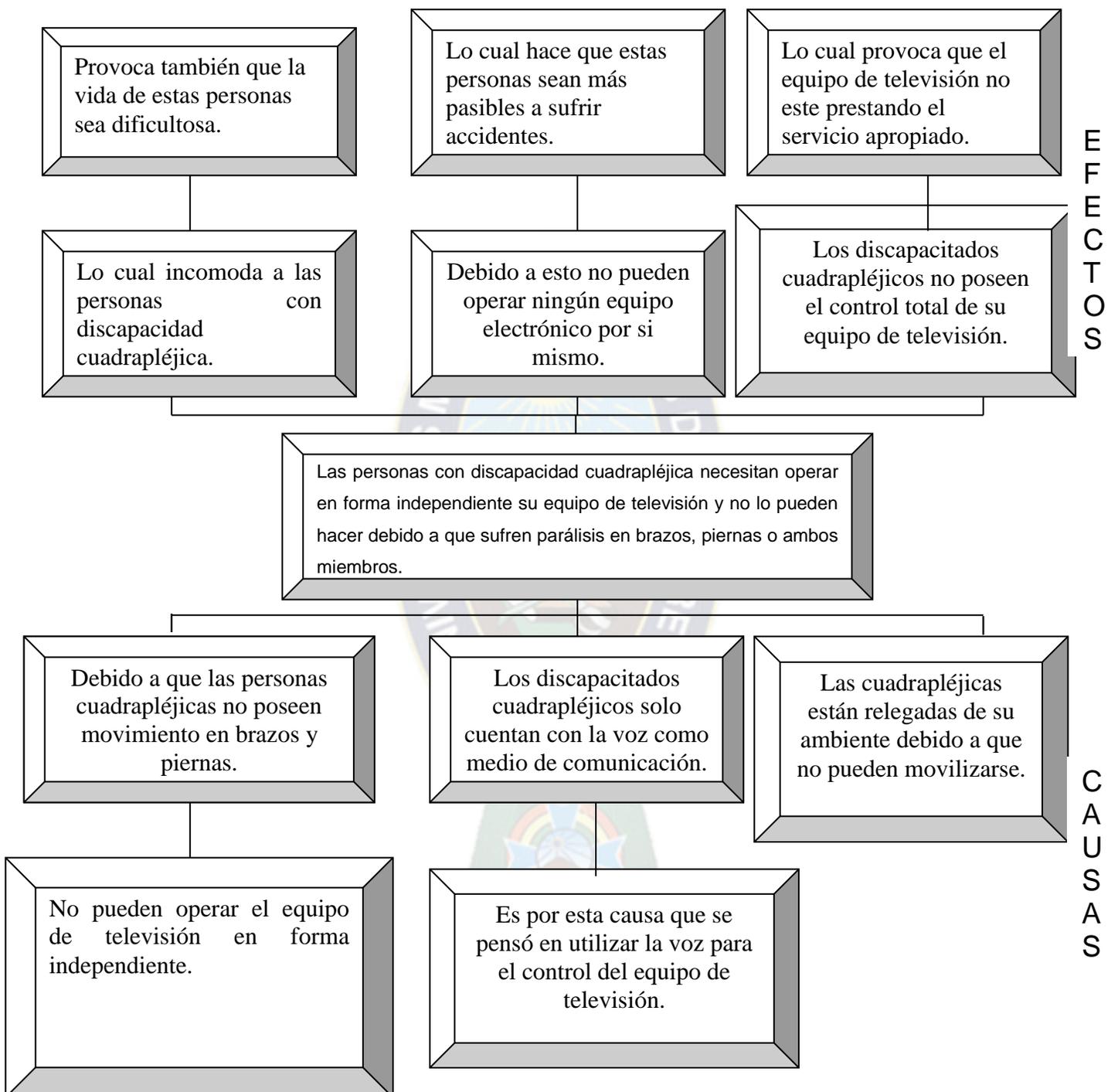
El reconocimiento de voz es un campo muy extenso para futuras investigaciones se recomiendan los siguientes temas:

- Reconocimiento de locutores
- Extensión del sistema a un mayor número de comandos y usuarios.
- Diferenciación de idiomas y acentos característicos de cada región.
- Identificación remota de voz utilizando redes de comunicaciones y pasarelas de datos y voz.
- Implementación de un sistema de reconocimiento de voz en un microprocesador o microcontrolador, lo cual abarataría enormemente los costos de hardware.
- Implementación de un sistema de comunicaciones inalámbrico entre software y hardware: Tecnología Bluetooth, radio frecuencia, etc.

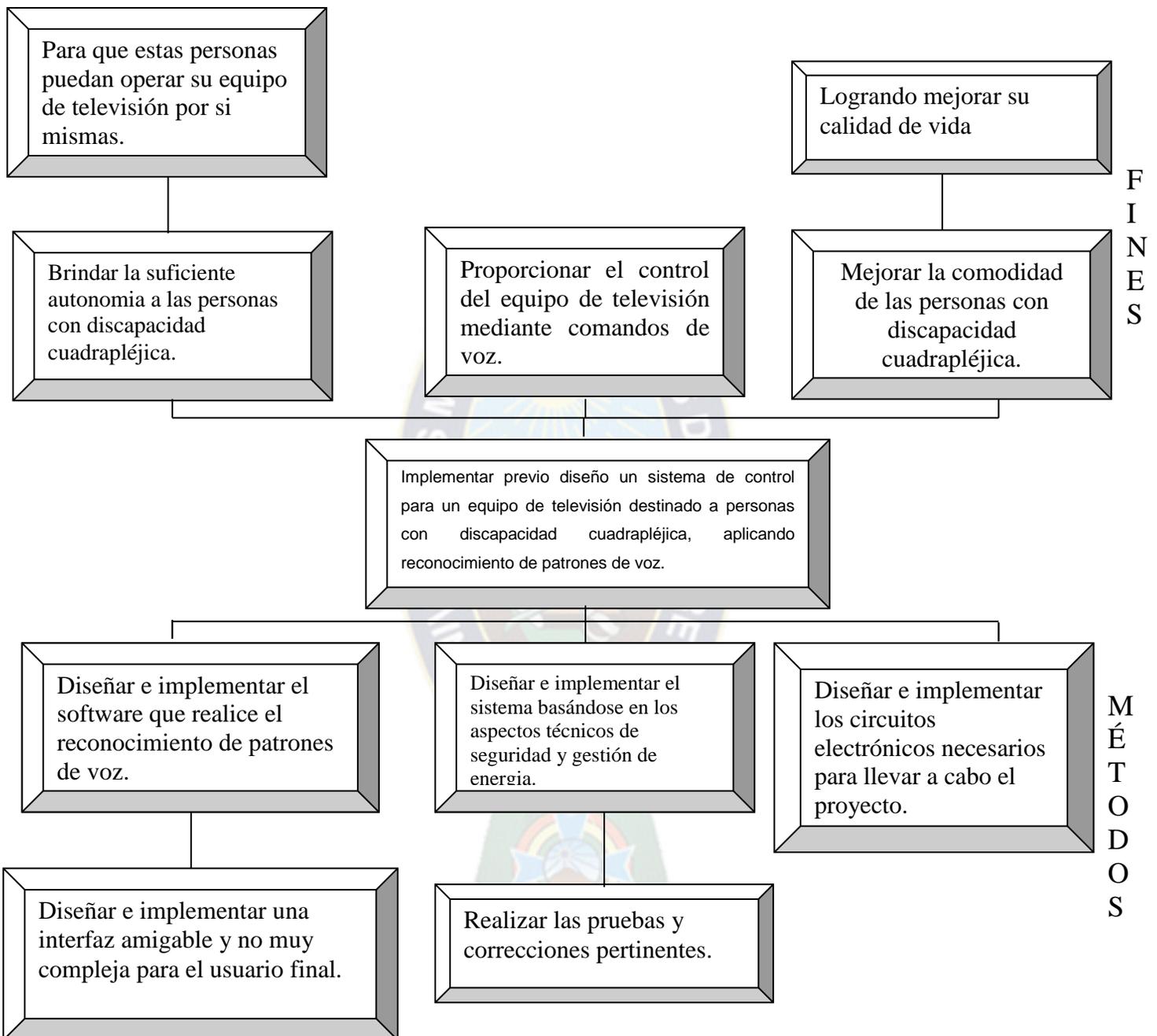


ANEXOS

ANEXO A ÁRBOL DE PROBLEMAS



ANEXO B ÁRBOL DE OBJETIVOS



ANEXO C FORMULARIO DE EVALUACIÓN DE PRUEBAS

#	METRICAS DE MC CALL	PREGUNTAS	PESIMO	MUY MALO	MALO	REGULAR	BUENO	MUY BUENO	EXCELENTE
			0	1	2	3	4	5	6
1	Compleción	¿El sistema cumple los requerimientos esperados?							
2	Consistencia	¿El diseño de las interfaces es el adecuado?							
3	Estandarización de datos	¿Los datos disponibles en las interfaces son los que se requieren?							
4	Tolerancia al error	¿Registra datos errados?							
5	Eficiencia de ejecución	¿El funcionamiento del sistema es el esperado?							
6	Generalidad	¿Son utilizados todos los componentes del sistema?							
7	Exactitud	¿Los resultados devueltos por el sistema son exactos?							
8	Estandarización de comunicaciones	¿Las respuestas enviadas por el puerto USB son rápidas?							
9	Instrumentación	¿Despliega mensajes de error cuando estos ocurren?							
10	Modularidad	¿Pueden utilizarse dos o más módulos a la vez?							
11	Operatividad	¿Es fácil la operatividad del sistema?							
12	Seguridad	¿La seguridad de acceso a los datos es adecuada?							
13	Formación	¿La utilización de las funciones del sistema son fáciles de asimilarlas?							

ANEXO D CARACTERISTICAS DEL PIC 18F4550

Configuración de pines

40-Pin PDIP

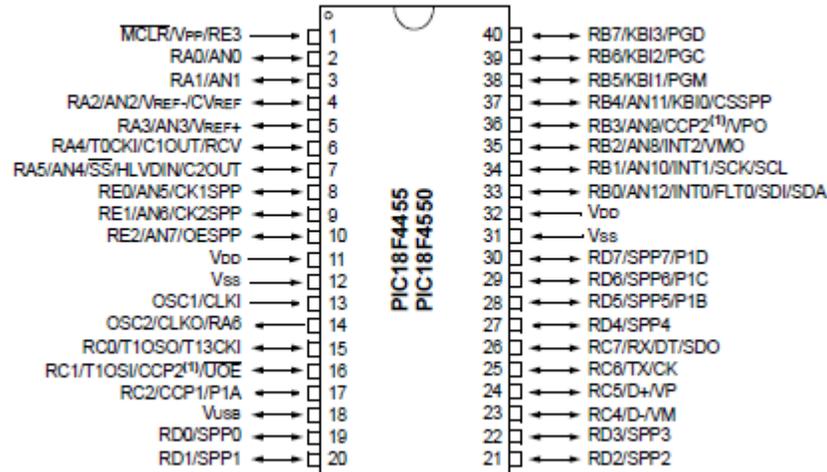
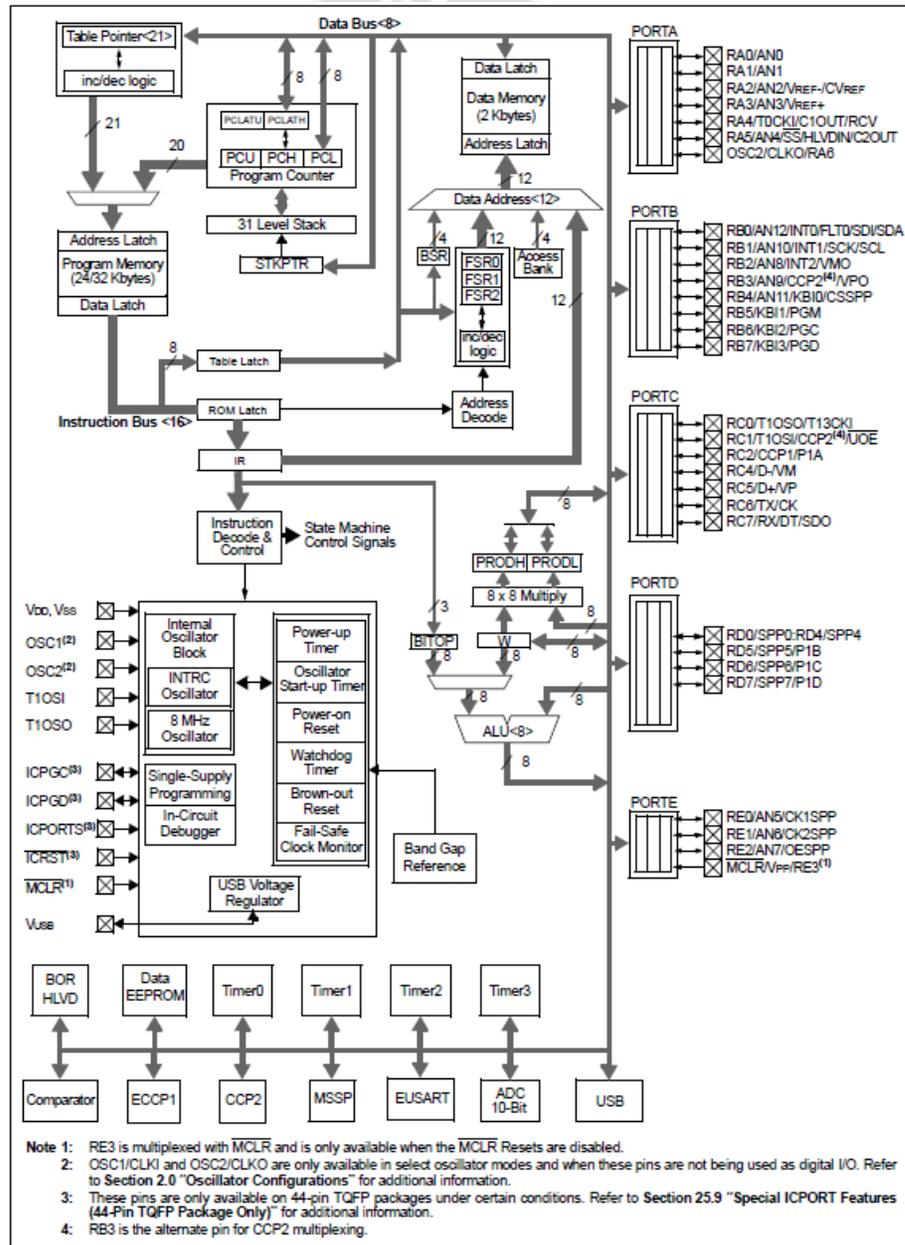


TABLE 1-1: DEVICE FEATURES

Features	PIC18F2455	PIC18F2550	PIC18F4455	PIC18F4550
Operating Frequency	DC – 48 MHz			
Program Memory (Bytes)	24576	32768	24576	32768
Program Memory (Instructions)	12288	16384	12288	16384
Data Memory (Bytes)	2048	2048	2048	2048
Data EEPROM Memory (Bytes)	256	256	256	256
Interrupt Sources	19	19	20	20
I/O Ports	Ports A, B, C, (E)	Ports A, B, C, (E)	Ports A, B, C, D, E	Ports A, B, C, D, E
Timers	4	4	4	4
Capture/Compare/PWM Modules	2	2	1	1
Enhanced Capture/Compare/PWM Modules	0	0	1	1
Serial Communications	MSSP, Enhanced USART	MSSP, Enhanced USART	MSSP, Enhanced USART	MSSP, Enhanced USART
Universal Serial Bus (USB) Module	1	1	1	1
Streaming Parallel Port (SPP)	No	No	Yes	Yes
10-Bit Analog-to-Digital Module	10 Input Channels	10 Input Channels	13 Input Channels	13 Input Channels
Comparators	2	2	2	2
Resets (and Delays)	POR, BOR, RESET Instruction, Stack Full, Stack Underflow (PWRT, OST), $\overline{\text{MCLR}}$ (optional), WDT	POR, BOR, RESET Instruction, Stack Full, Stack Underflow (PWRT, OST), $\overline{\text{MCLR}}$ (optional), WDT	POR, BOR, RESET Instruction, Stack Full, Stack Underflow (PWRT, OST), $\overline{\text{MCLR}}$ (optional), WDT	POR, BOR, RESET Instruction, Stack Full, Stack Underflow (PWRT, OST), $\overline{\text{MCLR}}$ (optional), WDT

Programmable Low-Voltage Detect	Yes	Yes	Yes	Yes
Programmable Brown-out Reset	Yes	Yes	Yes	Yes
Instruction Set	75 Instructions; 83 with Extended Instruction Set enabled			
Packages	28-Pin PDIP 28-Pin SOIC	28-Pin PDIP 28-Pin SOIC	40-Pin PDIP 44-Pin QFN 44-Pin TQFP	40-Pin PDIP 44-Pin QFN 44-Pin TQFP

Diagrama de bloques



REFERENCIAS BIBLIOGRÁFICAS

- El proceso unificado de desarrollo de software; Jacobson, Ivar Booch, gradyRumbaugh
- Ingeniería de software un enfoque practico; Pressman, Roger
- Metodología de la investigación científica; Zavala, Abel Andres
- Metodología de la investigación científica y técnicas de estudio; Uriarte Mora, Felipe.
- Inteligencia Artificial e Ingeniería del conocimiento; Pajares MartinSanz, Gonzalo Santos Peñas, Matilde
- Redes Neuronales Freeman & Skapura
- Procesamiento de Voz; Rocha Luis
- Fundamento de Señales y sistemas, usando la web y matlab; Kamen, Edgard W. Heck, Bonnie S.

Webgrafía

Ciclo de vida del software

- <http://img.redusers.com/imagenes/libros/lpcu097/capitulogratis.pdf>
- <http://alarcos.inf-cr.uclm.es/doc/ISOFTWAREI/Tema03.pdf>

Método científico

- <http://www.aulafacil.com/cursosenviados/Metodo-Cientifico.pdf>

Cocomo

- <http://creaweb.ei.uvigo.es/creaweb/Asignaturas/PPI/apuntes/cocomo2k.pdf>
- <http://www ldc.usb.ve/~teruel/ci4713/clases2001/cocomo2.html>
- <http://alarcos.inf-cr.uclm.es/DOC/PGSI/doc/teo/8/cocomo2-apuntes.pdf>

Modelo UML

- <http://es.tldp.org/Tutoriales/doc-modelado-sistemas-UML/doc-modelado-sistemas-uml.pdf>
- <http://elvex.ugr.es/decsai/java/pdf/3E-UML.pdf>
- <http://www.docirs.cl/uml.htm>
- <http://www.disca.upv.es/enheror/pdf/ActaUML.PDF>

Métodos de prueba: Caja Blanca y negra

- <http://www.udistrital.edu.co/comunidad/grupos/arquisoft/fileadmin/Estudiantes/Pruebas/HTML%20-%20Pruebas%20de%20software/node27.html>
- http://info_system.galeon.com/grupo8.html
- <http://www.ambysoft.com/essays/flootSpanish.html>

Redes Neuronales

- <http://electronica.com.mx/neural/informacion/backpropagation.html>

Voz humana

- fonatori <http://www.fceia.unr.edu.ar/acustica/biblio/fonatori.pdf>
- http://www.hezkuntza.ejgv.euskadi.net/r43573/es/contenidos/informacion/dia6_sigma/es_sigma/adjuntos/sigma_27/11_la_voz.pdf

La Transformada de Fourier

- art6 <http://www.emis.de/journals/DM/v5/art6.pdf>
- tdf2000 http://www.dm.uba.ar/materias/matematica_4/2007/1/tdf2000.pdf
- http://www.diac.upm.es/acceso_profesores/asignaturas/tdi/tdi/transformadas/pdf/four