



UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTA DE INGENIERÍA
CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

**“TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES”**

TRABAJO DE GRADUACION

TÍTULO DEL PROYECTO

**“CLASIFICACIÓN DE ESTADOS EMOCIONALES POR MEDIO DEL
RECONOCIMIENTO DE TONOS DE VOZ Y GRAMÁTICAS”**

AUTORES:

OLGA GUANGA

JUAN PABLO PIÑA

DIRECTORA: M. ENG. PAULINA VÉLEZ

Riobamba-Ecuador

AÑO 2016

INFORME DEL TUTOR

Yo, M. Eng. Paulina Vélez, en mi calidad de Tutor del trabajo de graduación titulado: “**CLASIFICACIÓN DE ESTADOS EMOCIONALES POR MEDIO DEL RECONOCIMIENTO DE TONOS DE VOZ Y GRAMÁTICAS**”, tengo a bien informar que el mencionado trabajo cumple con los requisitos exigidos para ser expuestos al público, luego de ser evaluado por el tribunal designado.

Riobamba, Agosto del 2016

Atentamente

A handwritten signature in blue ink, appearing to read 'Vélez Vélez', is written over a horizontal line.

M. Eng. Paulina Vélez

CERTIFICACIÓN

Los miembros del Tribunal de Graduación del proyecto de investigación de título: **CLASIFICACIÓN DE ESTADOS EMOCIONALES POR MEDIO DEL RECONOCIMIENTO DE TONOS DE VOZ Y GRAMÁTICAS** presentado por: Olga Guillermina Guanga Gualoto y Juan Pablo Piña Mayorga y dirigida por: M. Eng. Paulina Vélez.

Una vez escuchada la defensa oral y revisado el informe final del proyecto de investigación con fines de graduación escrito en la cual se ha constatado el cumplimiento de las observaciones realizadas, remite la presente para uso y custodia en la biblioteca de la Facultad de Ingeniería de la UNACH.

Para constancia de lo expuesto firman:

Ing. Fabián Gunsha

Presidente del Tribunal



Firma

M. Eng. Paulina Vélez

Director del Proyecto



Firma

Ing. Geovanny Cuzco

Miembro del Tribunal



Firma

AUTORÍA DE LA INVESTIGACIÓN

“La responsabilidad del contenido de este Proyecto de Graduación, corresponde exclusivamente a: Olga Guillermina Guanga Gualoto, Juan Pablo Piña Mayorga y de M. Eng. Paulina Vélez y el patrimonio intelectual de la misma a la Universidad Nacional de Chimborazo”

Autores:



Olga Guanga

060467948-0



Juan Pablo Piña

070549798-0

AGRADECIMIENTO

Agradezco en primer lugar a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

En especial a la M. Eng. Paulina Vélez, tutora de este trabajo de grado y quien ha puesto todo su compromiso en el desarrollo del mismo.

Olga Guanga

DEDICATORIA

Dedico este trabajo primeramente Dios, por haberme dado a vida y permitirme el haber llegado hasta este momento tan importante de mi formación profesional. A mis padres Hugo y Olga por ser el pilar más importante y por demostrarme siempre su cariño y apoyo incondicional para seguir adelante y no desmayar.

A mis hermanos por estar siempre presentes, acompañándome para poder realizar mis objetivos.

Olga Guanga

AGRADECIMIENTO

Agradezco primeramente a Dios por haberme orientado a lo largo de mis estudios, y brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

En especial a la M. Eng. Paulina Vélez, tutora de este trabajo de grado y quien ha puesto todo su compromiso en el desarrollo del mismo.

Juan Pablo Piña

DEDICATORIA

Esta tesis se la dedico a mi Dios quien supo guiarme por el buen camino, darme las fuerzas necesarias para seguir adelante y no desmayar.

A mi madre Petty Mayorga por inculcarme valores y principios de perseverancia en todo momento, a mis hermanas por su apoyo incondicional que me brindaron y en especial a mi hijo Juan Sebastián quien fue mi pilar fundamental en mi vida.

Juan Pablo Piña

ÍNDICE GENERAL

INFORME DEL TUTOR.....	II
CERTIFICACIÓN	III
AUTORÍA DE LA INVESTIGACIÓN	IV
ÍNDICE GENERAL.....	IX
ÍNDICE DE FIGURAS	XVI
ÍNDICE DE TABLAS	XVIII
RESUMEN.....	XIX
ABSTRACT.....	¡Error! Marcador no definido.
INTRODUCCIÓN	XXI
CAPÍTULO I.....	1
1 MARCO TEÓRICO	1
1.1 ANTECEDENTES	1
1.2 FUNDAMENTACIÓN TEÓRICA	3
1.2.1 LA VOZ HUMANA.....	3
1.2.1.1 CARACTERÍSTICAS DE LAS SEÑALES DE VOZ.....	3
1.2.1.2 ARTICULACIÓN	3
1.2.1.3 SEÑAL ACÚSTICA	5
1.2.1.3.1 FRECUENCIA Y AMPLITUD.....	5
1.2.1.3.2 RESONANCIA.....	6
1.2.1.3.3 ESTRUCTURA ARMÓNICA Y RUIDO	6
1.2.1.4 PERCEPCIÓN AUDITIVA	7
1.3 FRECUENCIAS DE LA VOZ HUMANA	7
1.3.1 EL TONO DE VOZ MASCULINA Y FEMENINA	8
1.3.2 CARACTERÍSTICAS DE LA VOZ HUMANA	8
1.3.2.1 TONO	8

1.3.2.2	TIMBRE	9
1.3.2.3	CANTIDAD	9
1.3.2.4	INTENSIDAD	9
1.3.3	CLASIFICACIÓN DE LOS SONIDOS DE LA VOZ.....	9
1.3.3.1	VOCALES Y CONSONANTES.....	10
1.3.3.2	ORALIDAD Y NASALIDAD	10
1.3.3.3	TONALIDAD	10
1.3.3.4	LUGAR Y MODO DE ARTICULACIÓN	10
1.3.3.5	SEGÚN LA POSICIÓN DE LOS ÓRGANOS ARTICULATORIOS	11
1.3.3.6	DURACIÓN	11
1.4	LAS EMOCIONES	12
1.4.1	LOS DIFERENTES TIPOS DE EMOCIONES	13
1.4.1.1	EMOCIONES BÁSICAS	13
1.4.1.2	EMOCIONES PRIMARIAS	13
1.4.1.3	EMOCIONES SECUNDARIAS	13
1.4.1.4	EMOCIONES NEGATIVAS	14
1.4.1.5	EMOCIONES POSITIVAS.....	14
1.4.1.6	EMOCIONES NEUTRAS.....	14
1.4.1.7	NATURALEZA DE LAS EMOCIONES.	14
1.4.2	LOS EFECTOS DE LAS EMOCIONES EN EL HABLA.	15
1.4.2.1	FRECUENCIA FUNDAMENTAL O PITCH.	15
1.4.2.2	DURACIÓN	15
1.4.2.3	CALIDAD DE VOZ.....	16
1.4.3	FUNCIÓN DE LAS EMOCIONES	16
1.4.3.1	FUNCIONES ADAPTATIVAS	16
1.4.3.2	FUNCIONES SOCIALES	17
1.4.3.3	FUNCIONES EMOCIONALES	17

1.4.4	TEORÍAS DE LAS EMOCIONES	18
1.4.4.1	CARACTERÍSTICAS DE LAS EMOCIONES	19
1.5	FUNCIONAMIENTO DE UN SISTEMA DE RECONOCIMIENTO DE VOZ	19
1.5.1.1	MÉTODOS Y ETAPAS PARA EL RECONOCIMIENTO DE VOZ ..	20
1.5.1.2	DIGITALIZACIÓN DE LA SEÑAL DE VOZ.....	21
1.5.1.3	CODIFICACIÓN DE LA SEÑAL DE VOZ.....	21
1.5.1.4	MUESTREO Y CUANTIZACIÓN.....	22
1.5.1.5	MUESTREO	22
1.5.1.6	CUANTIZACIÓN	23
1.5.1.7	CUANTIFICACIÓN UNIFORME	24
1.5.1.8	CUANTIFICACIÓN NO UNIFORME.....	24
1.5.1.9	CUANTIFICACIÓN LOGARÍTMICA	25
1.5.1.10	CUANTIFICACIÓN VECTORIAL.....	25
1.5.1.11	MEDICIÓN DE CARACTERÍSTICAS.....	26
1.5.1.12	CONVERSIÓN A/D O ADQUISICIÓN.....	26
1.5.1.13	OBTENCIÓN DE LA SEÑAL POR EL MICRÓFONO	26
1.5.1.14	FILTRACIÓN.....	26
1.5.1.15	ELIMINACIÓN DE RUIDO Y NORMALIZACIÓN	27
1.5.1.16	SEGMENTACIÓN	27
1.6	COEFICIENTES CEPSTRALES DE FRECUENCIA DE MEL	27
1.6.1	VENTANEOS.....	28
1.6.2	HAMMING	29
1.6.3	TRANSFORMADA DISCRETA DE FOURIER	29
1.6.4	FILTROS MEL.....	30
1.6.5	TRANSFORMADA COSENO DISCRETA.....	30
	CAPÍTULO II	32

2	METODOLOGÍA	32
2.1	TIPO DE ESTUDIO	32
2.1.1	DESCRIPTIVO	32
2.2	MÉTODOS, TÉCNICAS E INSTRUMENTOS	32
2.2.1	MÉTODOS	32
2.2.2	ANALÍTICO/DEDUCTIVO	32
2.2.3	TÉCNICAS	33
2.2.3.1	OBSERVACIÓN	33
2.2.3.2	INSTRUMENTACIÓN	33
2.3	POBLACIÓN MUESTRA	33
2.3.1	POBLACIÓN	33
2.3.2	MUESTRA	34
2.4	HIPÓTESIS	34
2.5	OPERACIONALIZACIÓN DE VARIABLES	34
2.6	PROCEDIMIENTOS	36
2.7	PROCESAMIENTO	37
2.7.1	DISEÑO	37
2.8	FUNCIONES UTILIZADAS	38
2.8.1.1.1	DESARROLLO	41
2.8.2	ENTRENAMIENTO Y BASE DE DATOS	44
2.8.2.1	CAPTURA DEL AUDIO	44
2.8.2.2	NORMALIZACIÓN	45
2.8.2.3	ELIMINACIÓN DE SILENCIOS	47
2.8.2.4	FRECUENCIA	48
2.8.2.5	MUESTREO	50
2.8.2.6	CUANTIZACIÓN	52
2.8.2.7	CODIFICACION	54

2.8.2.8	OBTENCION DE CARACTERISTICAS ESTADISTICAS	56
2.8.2.9	OBTENCION DE CARACTERISTICAS DE MEL.....	59
2.8.2.10	PROMEDIOS	61
2.8.2.11	BASE DE DATOS	62
2.8.3	ADQUISICIÓN	64
2.8.3.1	CAPTURA DEL AUDIO	64
2.8.3.2	NORMALIZACIÓN.....	65
2.8.3.3	ELIMINACIÓN DE SILENCIOS.....	66
2.8.3.4	FRECUENCIA	67
2.8.3.5	MUESTREO	69
2.8.3.6	CUANTIZACIÓN	71
2.8.3.7	CODIFICACIÓN.....	73
2.8.3.8	OBTENCION DE CARACTERISTICAS	74
2.8.3.9	OBTENCION DE CARACTERISTICAS DE MEL.....	77
2.8.3.10	PROCESO	80
2.9	COMPROBACIÓN DE LA HIPÓTESIS	82
2.10	PLANTEAMIENTO DE LA HIPÓTESIS ESTADÍSTICA	83
2.10.1	ESTABLECIMIENTO DEL NIVEL DE SIGNIFICANCIA.....	83
2.10.2	DETERMINACIÓN DEL VALOR ESTADÍSTICO DE PRUEBA. ...	83
2.11	PRUEBAS DE FUNCIONAMIENTO DEL SISTEMA.....	84
2.11.1	ESENARIO 1.....	84
2.11.2	ESCENARIO 2	85
2.11.3	ESCENARIO 3	85
2.11.4	ESCENARIO 4	86
2.11.5	ESCENARIO 5	87
2.11.6	ANÁLISIS DEL RESULTADO DEL GRUPO A	88
2.11.7	ANÁLISIS DEL RESULTADO DEL GRUPO B.....	89

2.11.8	PORCENTAJE DE RESULTADOS DE LOS GRUPOS DE PRUEBA .	89
2.11.9	VALOR ESTADÍSTICO DE LOS ESTADOS EMOCIONALES DEL GRUPO A	90
2.11.10	VALOR ESTADÍSTICO DE LOS ESTADOS EMOCIONALES DEL GRUPO B	90
2.11.11	CALCULO ESTADÍSTICO	91
2.11.12	VARIACIONES TOTALES DE LOS GRUPOS DE PRUEBA	92
	CAPÍTULO III	94
3	ANALISIS DE RESULTADOS	94
	CAPÍTULO IV	98
4	DISCUSIÓN	98
	CAPÍTULO V	99
5	CONCLUSIONES Y RECOMENDACIONES	99
5.1	CONCLUSIONES	99
5.2	RECOMENDACIONES	100
	CAPÍTULO VI	101
6	PROPUESTA	101
6.1	TITULO DE LA PROPUESTA	101
6.2	INTRODUCCIÓN	101
6.3	OBJETIVOS	101
6.3.1	OBJETIVO GENERAL	101
6.3.2	OBJETIVOS ESPECIFICOS	101
6.4	FUNDAMENTACIÓN CIENTIFICA	102
6.5	DESCRIPCIÓN DE LA PROPUESTA	102
6.6	DISEÑO ORGANIZACIONAL	103
6.7	MONITOREO Y EVALUACIÓN DE LA PROPUESTA	103

CAPÍTULO VII.....	104
7 BIBLIOGRAFÍA	104
CAPÍTULO VIII	106
8 ANEXOS	106

ÍNDICE DE FIGURAS

Figura 1. Corte esquemático del aparato fonatorio humano.	4
Figura 2. Funcionamiento de un sistema de Reconocimiento de Voz	20
Figura 3. Señal tratada Digitalmente para ser muestreada	22
Figura 4. Señal Cuantizada con 2 y 4 bits.	23
Figura 5. Señal Cuantizada Uniformemente	24
Figura 6. Señal Cuantizada No Uniformemente	25
Figura 7. Diagrama de la adquisición de Voz.....	26
Figura 8. Diagrama de eliminación de ruido y Normalización.....	27
Figura 9. Ventana de Hamming	29
Figura 10. Banco de filtros.....	30
Figura 11. Funcionamiento del software.....	36
Figura 12. Procesamiento	38
Figura 13. Audio Capturado.....	45
Figura 14. Normalización.....	46
Figura 15. Eliminación de silencios	48
Figura 16. Espectro de frecuencia	50
Figura 17. Espectro de frecuencia en dBs	50
Figura 18. Señal muestreada	52
Figura 19. Señal cuantizada	54
Figura 20. Señal codificada	55
Figura 21. Audio Capturado.....	64
Figura 22. Normalización.....	65
Figura 23. Eliminación de silencios	67
Figura 24. Espectro de frecuencia	69
Figura 25. Espectro de frecuencia en dBs	69
Figura 26. Señal muestreada	70
Figura 27. Señal cuantizada	73
Figura 28. Señal codificada	74
Figura 29. A) Señal en función del tiempo. B) Índice del canal en el tiempo. C) Índices cepstrales.....	80
Figura 30. Porcentaje de Pruebas del escenario1	84
Figura 31. Porcentaje de Pruebas del escenario2	85

Figura 32. Porcentaje de Pruebas del escenario 3	86
Figura 33. Porcentaje de Pruebas del escenario 4	87
Figura 34. Porcentaje de Pruebas del escenario 5	88
Figura 35. Estados emocionales del grupo A.....	90
Figura 36. Porcentajes de prueba	90
Figura 37. Resultado de Comparación.....	93
Figura 38. Señal de voz de entrada.	94
Figura 39. Eliminación de silencios	95
Figura 40. Señal de voz muestreada.....	95
Figura 41. Señal de voz cuantizada.	96
Figura 42. Señal codificada.	96
Figura 43. Espectro de frecuencia.	97
Figura 44. Esquema Organizacional	103

ÍNDICE DE TABLAS

Tabla 1. Funciones Adaptivas	17
Tabla 2. Operacionalización de Variables.....	35
Tabla 3. Diferencias de los escenarios de prueba.....	84
Tabla 4. Diferencias de los escenarios de prueba.....	85
Tabla 5. Diferencias de los escenarios de prueba.....	85
Tabla 6. Diferencias de los escenarios de prueba.....	86
Tabla 7. Diferencias de los escenarios de prueba.....	87
Tabla 8. Datos obtenidos en la aplicación del software a los estudiantes	88
Tabla 9. Datos obtenidos en la aplicación del software a los estudiantes del grupo B	89
Tabla 10. Porcentajes de prueba.....	89
Tabla 11. Variaciones de Estados Totales de los grupos de prueba.....	91
Tabla 12. Valor Final de prueba	92

RESUMEN

El procesamiento de señales es uno de los campos más importantes para el reconocimiento del habla y para reconocer las características del mismo, ya que mediante su análisis se justifican los conceptos en los que se fundamentan las técnicas de grabación, comportamientos acústicos, diseño de sistemas digitales y análogos. La señal que se analizará es la voz, como índice caracterizador de información del ser humano, cambios de amplitud, tono y timbre de un espectro vocal, pueden dar una descripción objetiva del comportamiento y las emociones humanas.

El presente proyecto describe el diseño de un software de clasificación de estados emocionales mediante el análisis de tonos de voz, para lo cual se han estudiado cuatro emociones (Miedo, Alegría, Tristeza, y Sorpresa), todo esto se realiza mediante la interacción hombre – máquina, ejecutando un análisis en la entrada al sistema, obteniendo características fundamentales de la señal de voz analizada en función del tiempo y frecuencia, que permiten posteriormente, una respuesta adecuada a través del diseño del software.



UNIVERSIDAD NACIONAL DE CHIMBORAZO

CENTRO DE IDIOMAS INSTITUCIONAL

Lic. Byron Soria

17 de Agosto del 2016

ABSTRACT

One of the most important fields to recognize speech is the signal processing and to explore the characteristics of it, because through its analysis the concepts on which recording techniques, acoustic behavior, design of digital and analog systems are based it is justified. The signal to be analyzed is the voice, as index information characterizing human, amplitude changes, rhythm and intonation of a voice spectrum, it can give an objective description of human behavior and emotions.

This study describes the design of a software classification of emotional states through analysis of voice tones, in the development of this project have been studied four emotions (Fear, Joy, Sadness and Surprise), this process is done by interaction man - machine executing an analysis on the input to the system, getting fundamental characteristics of the speech signal analyzed as a function of time and frequency, after it will allow an appropriate response through software design.



INTRODUCCIÓN

Los diferentes estados de emociones que presenta cada ser humano son esenciales en las actividades de interacción y cumplen un papel fundamental en la comunicación hablada, porque mediante el tono de voz se puede detectar el estado de ánimo de las personas.

Se plantea realizar una clasificación de los estados emocionales por medio del reconocimiento de tonos de voz a través del diseño de un software, que tiene como objetivo principal detectar las emociones de las personas para predecir los problemas de depresión. Este sistema funciona durante la interacción hombre-máquina, realizando ésta de forma natural.

La aplicación ha sido diseñada de tal forma que el usuario final pueda interactuar con la misma de forma sencilla, mediante un proceso de análisis y clasificación con algoritmos diseñados para clasificar los estados emocionales del usuario de forma rápida.

El reconocimiento automático de emociones se puede aplicarse por terapeutas como una herramienta de diagnóstico en medicina y psicología de manera inmediata. Los datos obtenidos por el reconocimiento de emociones por medio de la voz permiten recolectar la información necesaria, para determinar el estado emocional en tiempo real, por lo que puede ser empleado en centros de investigación públicos y privados en el mundo.

En la actualidad los sistemas de interacción humano- computadora son tecnologías que persiguen que la interacción sea bidireccional, en la cual la máquina escucha el mensaje del usuario y responde de manera instantánea puesto que se adapta al estado emocional de la persona de forma eficiente y amigable.

CAPÍTULO I

1 MARCO TEÓRICO

1.1 ANTECEDENTES

Para la búsqueda de información sobre la Clasificación de los Estados Emocionales utilizando tonos de voz y gramáticas se parte de la información de estudios ya existentes sobre el tema tales como:

“Reconocimiento de emociones a partir de voz basado en un modelo emocional continuo” En esta proyecto elaborado por Perez Espinosa & Reyes consiste en el reconocimiento de emociones a partir de la señal de voz enfocándose en bases de datos de emociones espontáneas, que realiza la interacción humano-computador. (Perez Espinosa & Reyes Garcia, 2010)

“Caracterización de voz empleando análisis tiempo-frecuencia aplicada al reconocimiento de emociones”.- El proyecto elaborado por Duque Sánchez & Morales Pérez, en este trabajo de tesis detalla una metodología para la caracterización de la señal de voz aplicada en el reconocimiento de estados emocionales utilizando tecnicas que pueden dividirse en diveretes categorias: Transformadas Tiempo-Frecuencia y Analisis Parametrico.(Duque Sánchez & Morales Pérez, 2009)

“Identificación automática del contenido afectivo de un texto y su papel en la presentación de información”.- El trabajo realizado por Virginia Francisco Gilmartín se encarga de clasificar palabras, textos o documentos de acuerdo a la opinión, emoción o sentimiento que expresan la aplicación utilizada es reconecedor de dialogo que existe entre máquina y persona (Virginia Francisco Gilmartín, 2008)

“Reconocimiento afectivo automático mediante el análisis de parámetros acústicos y lingüísticos del habla espontanea”.- El presente trabajo elaborado por Santiago Planet García determina el reconocimiento automático de emociones espontáneas basados en el análisis de la señal de voz mediante varias relaciones de síntesis efectiva relacionado con el análisis de la motivación. (Santiago Planet García, 2010)

Cabe indicar que no existen tesis o trabajos documentales ni escritos en la biblioteca de la institución beneficiaria del proyecto sobre temas relacionados con la presente propuesta de investigación cuyo nombre es: “CLASIFICACIÓN DE ESTADOS EMOCIONALES POR MEDIO DEL RECONOCIMIENTO DE TONOS DE VOZ Y GRAMÁTICAS”

A nivel global el desarrollo de sistemas de reconocimiento de voz se remonta a los años 1950 donde muchas de las técnicas utilizadas exitosamente tuvieron que esperar cerca de 10 años para poder pasar de la teoría a la práctica. Entre los primeros se mencionan los logrados por los laboratorios BELL en los años 1952, cuando se logró el reconocimiento aislado de dígitos para luego en 1959 lograr el reconocimiento de vocales y algunas consonantes. Estas tecnologías han ido evolucionando en todo el mundo con el pasar de los años hasta llegar inclusive a dispositivos móviles.

En américa latina el reconocimiento de emociones sigue experimentando grandes resultados tomando como referencia algunos estudios ya realizados para la mejora de sistemas de reconocimiento de las emociones de las personas por medio de la voz.

En ecuador se han desarrollado proyectos sobre reconocimiento de voz, mas no sobre la clasificación de estados emocionales por medio del reconocimiento de tonos de voz y gramáticas con lo que se logra detectar los estados emocionales de las personas y de esta manera hacer posible que la interacción humano-máquina se asemeje a la comunicación humana.

La institución que se beneficia del presente proyecto es la UNIVERSIDAD NACIONAL DE CHIMBORAZO ya que se plantea una alternativa del estudio y análisis de las emociones por medio de la voz.

Es por ello que se aceptó el desafío de brindar alternativas de reconocimiento de reconocimiento de voz.

1.2 FUNDAMENTACIÓN TEÓRICA

1.2.1 LA VOZ HUMANA

La voz humana, es la representación de un ser, de su pensar y sentir, es su sello distintivo, “es una herramienta vital en el proceso de comunicación con otros seres, por medio de ella se expresan sentimientos y conocimientos, se imparten instrucciones como a su vez las recibe, también es utilizada para transmitir emociones”. (Farfán, 2013)

La voz humana cambia normalmente en tres etapas de nuestra vida, desde niños, adolescentes y en la vejez, también puede variar por factores externos como alteraciones físicas en alguno de los órganos que la produce, o por causa de algún accidente.

1.2.1.1 CARACTERÍSTICAS DE LAS SEÑALES DE VOZ

Los sistemas de reconocimiento de voz se enfocan en las palabras y los sonidos que distinguen una palabra de la otra en un idioma.

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

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

Los tres enfoques proveen ideas y herramientas para obtener mejores y más eficientes resultados en el reconocimiento. (Pelton, 1992)

1.2.1.2 ARTICULACIÓN

La articulación se compone del sistema fonador, en el cual la voz humana es producida, está formado por los pulmones que proporcionan un flujo de aire, la laringe, que contiene las cuerdas vocales, la faringe, las cavidades oral (o bucal) y

nasal y una sucesión de elementos articulatorios: los labios, los dientes, el alvéolo, el paladar, el velo del paladar y la lengua (figura 1). (Vives, 1993)

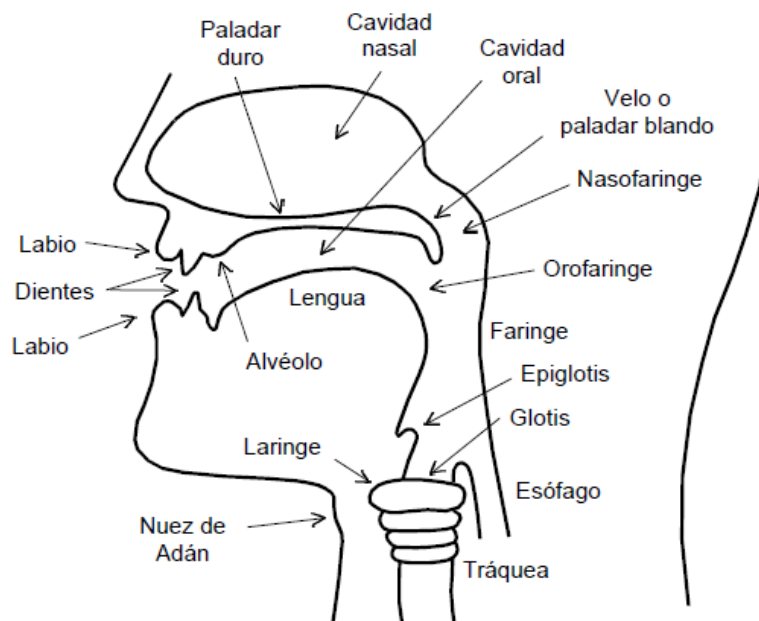


Figura 1. Corte esquemático del aparato fonatorio humano.

Fuente. Federico Miyara

Tráquea.- Esta situada antes del esófago, se extiende entre la laringe y los bronquios principales derecho e izquierdo la función es de conducir el aire hacia los pulmones o fuera de ellos.

Laringe.- Su función es proteger las vías respiratorias y producir los sonidos bajo la acción del aire espiratorio.

Los despliegues vocales varían dependiendo el sexo en el hombre crecen entre 4 y 11 mm y en la mujer entre 1,5 y 4 mm, el cambio de voz masculina es más evidente. Los cambios que se dan antes de la pubertad de los pliegues vocales el hombre y la mujer son parecidos, varían después de la pubertad en el hombre que tiene casi el doble de longitud en los pliegues vocales por término medio de las mujeres. Durante la pubertad la voz femenina desciende alrededor de 2,5 semitonos y la voz masculina alrededor de una octava la voz durante la pubertad. (Daniel, 1988)

Boca.- Los pliegues vocales adoptan su forma y volumen siendo la boca el principal resonador de la voz, depende de los movimientos o cambios de la posición de la lengua, los labios, el velo y la mandíbula. (Daniel, 1988)

Mediante el sonido generado por los pliegues vocales es impartida por la boca la intensidad o volumen final del sonido será directamente proporcional al área de abertura de este.

Labios.- Limitada la hendidura bucal (fisura oral). Lateralmente se unen formando la comisura labial y el ángulo de la boca. (Daniel, 1988)

Lengua.- Es un órgano musculoso situado en el suelo de la boca que implica en la succión, la masticación, la deglución y la fonación. (Daniel, 1988)

1.2.1.3 SEÑAL ACÚSTICA

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

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

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

Algunas características importantes del análisis acústico son: (Harriton, 2005)

1.2.1.3.1 FRECUENCIA Y AMPLITUD

Todos los sonidos causan movimientos entre las moléculas del aire. Algunos sonidos, tales como los que produce una cuerda de guitarra, producen patrones regulares y prolongados de movimiento del aire. Los patrones de sonidos más simples son los sonidos puros (pure tones), y se pueden representar gráficamente por una onda sinusoidal. (Harriton, 2005)

Frecuencia: Número de vibraciones (ciclos) del tono por segundo, 100 por segundo

- ciclos/segundo = 100 Hz.
- Tonos altos = Mayor frecuencia

- Tonos bajos = Menor frecuencia

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

1.2.1.3.2 RESONANCIA

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

La mayoría de los sonidos incluyendo del habla tienen una frecuencia dominante llamada frecuencia fundamental también conocida como pitch (tono) que se combina con frecuencias secundarias en el habla, la frecuencia fundamental es la velocidad a la que vibran las cuerdas vocales al producir un fonema sonoro. (Harriton, 2005)

1.2.1.3.3 ESTRUCTURA ARMÓNICA Y RUIDO

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

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

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

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

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

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

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

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

1.2.1.4 PERCEPCIÓN AUDITIVA

Es un sistema basado en un enfoque del conocimiento de la percepción humana se sabe que el sistema auditivo está adaptado a la percepción de la voz.

La frecuencia que detecta el oído humano está entre 20Hz a 20,000Hz y es más sensible entre 1000 y 6000 Hz también es sensible a los cambios pequeños en frecuencia ancho de banda crítico para el habla. (Faúndez Zanuy, 2000)

1.3 FRECUENCIAS DE LA VOZ HUMANA

La frecuencia de la voz humana es el número de vibraciones por segundo, es medida en Hercios. La frecuencia vibratoria de las cuerdas vocales es utilizada para medir el tono de la voz de las personas. Cuando la frecuencia es muy alta, la voz se vuelve más aguda, y, cuando la frecuencia es más baja, la voz se vuelve más grave. Según varios autores, se puede delimitar la frecuencia de la voz masculina entre 50 y 200 hercios y la voz femenina entre 150 y 350 hercios. (Miyara, 2005)

Es posible establecer el tono medio de voz de las personas, según varios autores, es posible fijar que en las voces masculinas el tono medio es de 125 hercios y en las voces femeninas el tono medio es de 215 hercios.

1.3.1 EL TONO DE VOZ MASCULINA Y FEMENINA

Se entiende por tono a la altura o elevación de la voz que se obtiene de la frecuencia de la vibración de las cuerdas vocales. Estos músculos al vibrar muchas veces por segundo, aumentan su tensión, de tal modo que cuando la altura es mayor, la voz se eleva y, se escucha una voz más aguda. De tal modo, al disminuir la frecuencia, se reduce las vibraciones por segundo, lo cual produce menor tensión de las cuerdas vocales y, se puede escuchar una voz más grave. Las vibraciones de un sonido agudo, en la unidad de tiempo, son más numerosas que las vibraciones de un sonido grave. La escala de frecuencia tonal es la que permite clasificar los sonidos mediante el tono, que es la propiedad de la voz, esta escala va de más agudo a más grave. (Antón, 2001)

El tono depende de la longitud que presenten los pliegues vocales, cuanto mayor sea, más grave será la voz. Al contrario, cuanto más cortas sean las cuerdas, las voces serán más agudas. (Antón, 2001).

Por lo general los pliegues de un niño miden entre cinco y doce milímetros, por lo que la voz de un es más aguda que la de una mujer, ya que tienen pliegues que miden entre catorce y dieciocho milímetros, ésta más aguda que la de un hombre, ya que tienen pliegues con longitudes entre dieciocho y veinticinco milímetros.

1.3.2 CARACTERISTICAS DE LA VOZ HUMANA

1.3.2.1 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. (M, 2008)

1.3.2.2 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. Por el timbre se reconoce a la persona que habla, aun cuando no se la perciba. Hay voces bien timbradas y agradables, más las hay también blancas, roncadas y chillonas.

1.3.2.3 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.

1.3.2.4 INTENSIDAD

Es la cantidad de energía acústica que se irradia desde la boca de un hablante, es decir la presión sonora que ejerce ese flujo aéreo.

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. (A., 2007)

1.3.3 CLASIFICACIÓN DE LOS SONIDOS DE LA VOZ

El aparato fonatorio pueden clasificarse en diferentes aspectos del fenómeno de emisión: (M, 2008)

- a) Según su carácter vocálico o consonántico.
- b) Según su oralidad o nasalidad
- c) Según su carácter tonal (sonoro) o no tonal (sordo)
- d) Según el lugar de articulación
- e) Según el modo de articulación
- f) Según la posición de los órganos articulatorios

g) Según la duración

1.3.3.1 VOCALES Y CONSONANTES

Los sonidos se dividen tradicionalmente en vocálicos y consonánticos cuando se pronuncia una consonante las cuerdas vocales se mantienen muy tensas, vibran con una frecuencia relativamente más baja que cuando se pronuncia una vocal, y por lo tanto, la consonante tiene un tono más bajo.

1.3.3.2 ORALIDAD Y NASALIDAD

Los fonemas en los que el aire pasa por la cavidad nasal se denominan nasales, en tanto que aquéllos en los que sale por la boca se denominan orales.

1.3.3.3 TONALIDAD

La tonalidad de voz hace referencia al volumen, proyección y tono de tu voz, es la cualidad subjetiva de la onda sonora asociada más estrechamente a la frecuencia, la tonalidad se define como el aspecto de la sensación auditiva en términos desigual, los sonidos pueden ser ordenados en una escala que va desde “graves” hasta “agudos”.

1.3.3.4 LUGAR Y MODO DE ARTICULACIÓN

La articulación es la Posición y movimiento de los órganos de la voz para la pronunciación de una vocal o consonante. (Miyara, 2005)

En función del lugar o punto de articulación se tienen fonemas:

- Bilabiales: Oposición de ambos labios
- Labiodentales: Oposición de los dientes superiores con el labio inferior
- Linguodentales: Oposición de la punta de la lengua con los dientes superiores
- Alveolares: Oposición de la punta de la lengua con la región alveolar

- Palatales: Oposición de la lengua con el paladar duro
- Velares: Oposición de la parte posterior de la lengua con el paladar blando.
- Glotales: Articulación en la propia glotis

A su vez, para cada punto de articulación ésta puede efectuarse de diferentes modos, dando lugar a fonemas:

- Oclusivos: la salida del aire se cierra momentáneamente por completo
- Fricativos: el aire sale atravesando un espacio estrecho
- Africados: oclusión seguida por fricación
- Laterales: la lengua obstruye el centro de la boca y el aire sale por los lados
- Vibrantes: la lengua vibra cerrando el paso del aire intermitentemente
- Aproximantes: La obstrucción muy estrecha que no llega a producir turbulencia.

1.3.3.5 SEGÚN LA POSICIÓN DE LOS ÓRGANOS ARTICULATORIOS

Los órganos intervienen en su producción, la posición que éstos son posiciones que hacen variar la trayectoria del aire en su viaje por la garganta hasta que es expulsado por la boca o la nariz, produciendo sonidos diferentes.

1.3.3.6 DURACIÓN

La duración es el tiempo durante el cual se mantiene un sonido, está determinada por la longitud, que indica el tamaño de una onda, que es la distancia entre el principio y el final de una onda completa (ciclo); es decir que por duración de los sonidos pueden ser largos o cortos. (Daniel, 1988)

1.4 LAS EMOCIONES

Las emociones son reacciones a las informaciones (conocimiento) que se obtiene en nuestras relaciones con el entorno. La intensidad de la reacción está en función de las evaluaciones subjetivas que se realiza sobre cómo la información recibida va a afectar nuestro bienestar. En estas evaluaciones subjetivas intervienen conocimientos previos, creencias, objetivos personales, percepción de ambiente provocativo, etc. Una emoción depende de lo que es importante para una persona. Si la emoción es muy intensa puede producir disfunciones intelectuales o trastornos emocionales (fobia, estrés, depresión). (Vivas, 2007)

- En el ser humano, la experiencia de una emoción involucra tres componentes: Fisiológico, verbal y conductual. La alegría, la ira o la ansiedad son ejemplos de emociones.
- En la emoción un conjunto de cogniciones, actitudes y creencias sobre el mundo, son utilizadas para valorar una situación concreta e influyen en el modo en el que se percibe dicha situación.

La voz es el principal modo de comunicación entre las personas y por esta causa constantemente se ha estudiado los elementos que funcionan en la producción de voz humana y se han creado gran cantidad de sistemas capaces de simular y reconocer voz electrónicamente.

Las emociones no es un fenómeno simple, sino que muchos factores afectan e intervienen en las mismas, ya que estas son experimentadas cuando ocurre un suceso inesperado, por tal la emoción afecta a los movimientos musculares del aparato respiratorio y la laringe y ello modifica el tono de voz de las personas. Así, se observa que, al hablar, las palabras no son emitidas desnudas sino que van acompañadas de nuestras emociones. (Salías, 2005)

1.4.1 LOS DIFERENTES TIPOS DE EMOCIONES

Una de las clasificaciones más empleadas es la de distinguir entre emociones básicas y emociones complejas o secundarias. Las emociones básicas, también denominadas primarias, puras o elementales, se reconocen por una expresión facial característica. (Ekma, 1992)

1.4.1.1 EMOCIONES BÁSICAS

Inician con rapidez y duran unos segundos cada vez, por lo general se reconocen como tales: la alegría, la aflicción o tristeza, la ira, el miedo, la sorpresa y el asco. No son aprendidas sino que forman parte de la configuración del ser humano. Esto se evidencia en la presencia de las mismas expresiones faciales en los ciegos congénitos. Se pueden usar palabras o conceptos distintos para describir una determinada emoción, pero la expresión facial es la misma. (Ekman, 1992)

1.4.1.2 EMOCIONES PRIMARIAS

Poseen una alta carga genética, en el sentido que presentan respuestas emocionales pres organizados que, aunque son modeladas por el aprendizaje y la experiencia, están presentes en todas las personas y culturas. (Ekma, 1992)

1.4.1.3 EMOCIONES SECUNDARIAS

Proceden de las primarias, se deben en gran parte al desarrollo individual y sus respuestas difieren ampliamente de unas personas a otras. (Ekma, 1992)

1.4.1.4 EMOCIONES NEGATIVAS

Implican sentimientos desagradables, valoración de la situación como dañina y la movilización de muchos recursos para su afrontamiento. Ej.: el miedo, la ira, la tristeza y el asco. (Ekma, 1992)

1.4.1.5 EMOCIONES POSITIVAS

Son aquellas que implican sentimientos agradables, valoración de la situación como beneficiosa, tienen una duración temporal muy corta y movilizan escasos recursos para su afrontamiento. Ej.: felicidad (Ekma, 1992)

1.4.1.6 EMOCIONES NEUTRAS

Son las que no producen intrínsecamente reacciones ni agradables ni desagradables, es decir que no pueden considerarse ni como positivas ni como negativas, tienen como finalidad el facilitar la aparición de posteriores estados emocionales. (Ekma, 1992)

1.4.1.7 NATURALEZA DE LAS EMOCIONES.

La emoción es una serie de fenómenos complejos que se ven afectados por un gran número de factores.

Para Scherer la emoción se puede describirse también como la interfaz del organismo con el mundo exterior, marcando tres funciones principales de las emociones: (Vivas, 2007)

- a) Muestran la apreciación de la importancia de un estímulo.
- b) Preparan fisiológica y físicamente al organismo para la acción apropiada.
- c) Comunican el estado del organismo y sus intenciones de comportamiento a otros organismos que le rodean.

1.4.2 LOS EFECTOS DE LAS EMOCIONES EN EL HABLA.

Darwin describió en su libro “La expresión de las emociones en el hombre y los animales”, sobre cómo afectan las emociones en el comportamiento y el lenguaje. Actualmente los estudios acerca de las emociones en el habla se realiza por investigadores que analizan la señal de la voz de las personas, gracias a estos estudios se ha logrado identificar muchos de los elementos del habla en las personas que son utilizados para la expresión de las emociones, tales como: (REEVE, 2002)

- Frecuencia fundamental.
- Calidad de Voz.
- Duración.

1.4.2.1 FRECUENCIA FUNDAMENTAL O PITCH.

El “pitch” o frecuencia fundamental F0, es la frecuencia a la que se realiza la vibración de las cuerdas vocales. Las características de la frecuencia fundamental son consideradas como una de las principales portadoras de información sobre las emociones en la voz de las personas. Una frecuencia fundamental elevada indica un mayor grado de excitación. (REEVE, 2002)

1.4.2.2 DURACIÓN

La duración tiene una relación directa con la velocidad del habla, sus efectos son el ritmo y la velocidad. El ritmo hace referencia al acento que refleja las características y las experiencias especiales del individuo. Se debe tomar en cuenta el número de palabras por minuto lo cual aumenta o reduce el tiempo de duración del habla, además del número de pausas y la duración de las mismas. (REEVE, 2002)

1.4.2.3 CALIDAD DE VOZ

La calidad de voz hace referencia a la claridad con la que se habla, la intensidad, y factores externos que pueden intervenir en la calidad, se debe tomar en consideración que cada persona tiene su propio tono de voz. (REEVE, 2002)

1.4.3 FUNCIÓN DE LAS EMOCIONES

Todas las emociones tienen alguna función que las hace útiles y permite que las personas ejecuten con eficacia las reacciones conductuales apropiadas, independientemente de la cualidad placentera que puedan generar; pues incluso las emociones más desagradables tienen funciones importantes en la adaptación social y el ajuste personal. (REEVE, 2002)

1.4.3.1 FUNCIONES ADAPTATIVAS

Tal vez una de las funciones más importantes de la emoción es la de preparar a las personas para que ejecuten eficazmente los comportamientos requeridos por las condiciones ambientales y sociales en las que se encuentren, haciendo que movilicen a la energía necesaria para ello y guiándolo (acercando o alejando) hacia un objeto determinado.

Plutchik (1980) considera ocho funciones principales de las emociones y ayuda a construir un lenguaje funcional que identifique cada una de dichas reacciones con la tarea adaptativa que le corresponde. La correspondencia entre la emoción y su función se enuncia en el siguiente cuadro:

Lenguaje subjetivo	Lenguaje funcional
Miedo	Protección
Alegría	Reproducción
Tristeza	Reintegración
Sorpresa	Exploración

Tabla 1. Funciones Adaptivas

Fuente. Plutchik

1.4.3.2 FUNCIONES SOCIALES

La función principal de las emociones es facilitar la aparición de los comportamientos apropiadas; así, la expresión de las emociones permite a los demás predecir las conductas relacionadas con ellas, dándole un importante valor en las relaciones interpersonales como la de facilitar la interacción social, controlar el comportamiento de otras personas, permitir la comunicación o expresión de los estados afectivos y promover el comportamiento pro-social. (REEVE, 2002)

Emociones como la felicidad favorecen los vínculos sociales y relaciones interpersonales en diferentes ámbitos como con los recién nacidos, en las relaciones de pareja, con los amigos, entre otros: la ira puede generar respuestas de evitación, de confrontación e incluso de agresión. De todas formas la posibilidad de expresar las emociones puede considerarse una posibilidad con la que cuentan las personas que promueven la realización de comportamientos apropiados por parte de los demás.

1.4.3.3 FUNCIONES EMOCIONALES

La relación entre emoción y motivación es íntima, ya que se trata de una experiencia presente en cualquier tipo de actividad que posee las dos principales características de la conducta motivada, dirección e intensidad. La emoción energiza la conducta motivada. Una conducta que tenga carga emocional probablemente se realizará de

forma más vigorosa, con más empeño. La ira, por ejemplo, facilita las reacciones defensivas, la alegría la reacción interpersonal, la sorpresa la atención ante nuevos estímulos. Cuando la emoción dirige u orienta el comportamiento o la evitación del objetivo de la conducta motivada en función de las características placenteras o displacenteras de la emoción. (REEVE, 2002)

1.4.4 TEORÍAS DE LAS EMOCIONES

Estas teorías se basan en la fisiología, las cogniciones y la interacción de los factores físicos mentales: (Alzina, 2003)

Teoría de James (1884) y Lange(1885).

La emoción tiene lugar cuando el individuo interpreta sus respuestas fisiológicas o sensaciones físicas los estímulos que provocan la emoción, tales como el aumento de nuestro ritmo cardiaco y los espasmos musculares, esto conlleva que cada emoción diferente debe tener su propio patrón fisiológico.

Teoría de Cannon (1927) y Bard (1938).

La emoción es fundamentalmente un acontecimiento cognitivo. Todas las reacciones físicas son iguales para diferentes emociones. Sobre la base (únicamente) de las señales fisiológicas no es posible distinguir una emoción de otra.

Teoría de Sachachter-Singer(1962).

Esta teoría sostiene que la emoción es debida a la evaluación cognitiva del acontecimiento y a la respuesta corporal. La persona nota los cambios fisiológicos, advierte lo que ocurre a su alrededor y denomina sus emociones de acuerdo a ambos tipos de observación, la intensidad de la emoción esta fundamentalmente determinada por los cambios fisiológicos.

1.4.4.1 CARACTERÍSTICAS DE LAS EMOCIONES

Éstas son algunas de las características de las emociones: (REEVE, 2002)

1. Las emociones son un conjunto complejo de respuestas neurales y químicas que conforman un patrón. Todas cumplen un papel regulador destinado a crear situaciones de ventaja para el organismo que presenta el fenómeno.
2. Las emociones son procesos determinados biológicamente de manera innata.
3. Los mecanismos que producen las emociones utilizan un conjunto restringido de regiones subcorticales del cerebro del tronco del encéfalo, hipotálamo, cerebro anterior basal y amígdala.
4. Los mecanismos de la emoción se desencadenan automáticamente, sin deliberación consciente. Las emociones tienen un importante componente de estereotipia y automatismo a pesar de las variaciones culturales.
5. Las emociones se dan en el cuerpo del organismo (medio interno, sistemas visceral, vestibular y músculo-esquelético) y también afectan el funcionamiento de variados circuitos cerebrales.

1.5 FUNCIONAMIENTO DE UN SISTEMA DE RECONOCIMIENTO DE VOZ

El reconocimiento de voz generalmente es utilizado como una interfaz entre humano y computadora para algún software.

Debe cumplir 3 tareas:

- **Pre-procesamiento:** Convierte la entrada de voz a una forma que el reconocedor pueda procesar.
- **Reconocimiento:** Identifica lo que se dijo.
- **Comunicación:** Envía lo reconocido al sistema (Software/Hardware) que lo requiere.

Componentes en una aplicación:

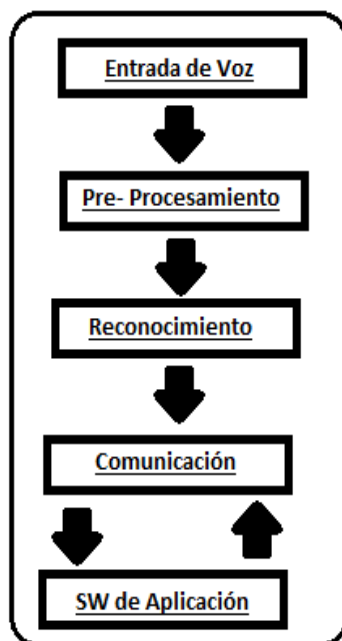


Figura 2. Funcionamiento de un sistema de Reconocimiento de Voz

Fuente. Autores

La comunicación que existen es bilateral, cuando se trata de una comunicación bilateral la interfaz de voz está relacionado con otras aplicaciones el reconocedor permitiendo específicamente que palabras o estructuras del sistema puede utilizar.

Los procesos de pre-procesamiento, reconocimiento y comunicación deberían ser invisibles al usuario de la interfaz. El usuario lo nota en el desempeño del sistema de manera indirecta como: certeza en el reconocimiento y velocidad. Estas características las utiliza para evaluar una interfaz de reconocimiento de voz. (Pereira, 2003)

1.5.1.1 MÉTODOS Y ETAPAS PARA EL RECONOCIMIENTO DE VOZ

Los diferentes tipos de Métodos que existen para el reconocimiento de Voz, así como las Etapas por las cuales debe pasar un reconocedor de voz para poder procesar la voz y manejarla por medio de software.

1.5.1.2 DIGITALIZACIÓN DE LA SEÑAL DE VOZ

Para poder manipular la señal de voz este debe de ser transformada a una señal digital para poder ser utilizada por el software de programación, este proceso se lleva a cabo en etapas con consideraciones variables que hacen que la adquisición de la señal sea de forma correcta. (REEVE, 2002)

1.5.1.3 CODIFICACIÓN DE LA SEÑAL DE VOZ

La codificación de una señal se refiere al método para convertir digitalmente el habla para utilizarlo en diversos ambientes, desde juguetes parlantes, CD's hasta transmisiones vía telefónica. (Faúndez, 2000)

Para utilizar la voz como dato en aplicaciones tan diversas como el voice mail, anotaciones en un texto o un directorio parlante, es necesario almacenar la voz de manera que una computadora pueda recuperarla.

La presentación digital de la voz provee también con las bases para el reconocimiento y síntesis de voz.

Algunos factores que se deben considerar para codificar la señal dependiendo del uso que se le dé son:

- Memoria y Ancho de banda necesario para flexibilidad de uso.
- Costo de transmisión.
- Rangos de calidad
- Aplicación.
- Codificadores de Voz.
- Inteligibilidad: La mayor calidad posible
- Error e inteligibilidad.
- Edición Simple.
- Eliminación del silencio: Ahorrar espacio
- Time-scaling: Reproducción más rápida.
- Adaptación de Velocidad
- Robustez

1.5.1.4 MUESTREO Y CUANTIZACIÓN

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

Un micrófono convierte esas variaciones en presión del aire a variaciones en voltaje. Una señal Análoga se puede transmitir a través de un circuito telefónico (voltaje) o almacenados en una cinta magnética (flujo magnético). (Faúndez Zanuy, 2000)

En el Mundo Real los estímulos sensoriales son análogos. Sin embargo, para las computadoras es necesario digitalizar la señal (primera fase del procesamiento de la señal).

1.5.1.5 MUESTREO

El muestreo transforma una señal analógica o señal continua a una señal digital o discreta, es decir, de la señal original (analógica) se toman muestras cada determinado tiempo y se hace una copia de su valor en tiempos definidos, de esta forma surge la frecuencia de muestreo que es la cantidad de muestras que se tienen de una señal en una unidad de tiempo y se mide en Hz (ciclos por segundo). (Kai-Fu, 1990)

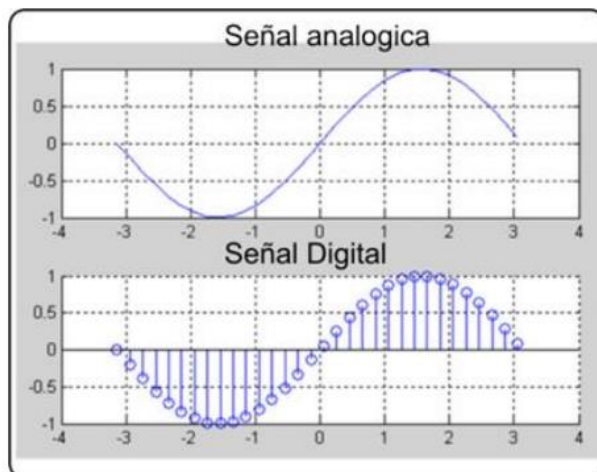


Figura 3. Señal tratada Digitalmente para ser muestreada

Fuente. Kai-Fu

El muestreo asigna un valor numérico a la señal en unidades discretas de tiempo constante dependiendo de la frecuencia Nyquist que especifica la frecuencia máxima a la que una señal puede reproducirse completamente. (Faúndez Zanuy, 2000)

El teorema de Nyquist establece que:

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

1.5.1.6 CUANTIZACIÓN

El proceso de cuantificación es uno de los pasos que se siguen para lograr la digitalización de una señal analógica. Este proceso es convertir una sucesión de muestras de amplitud continua en una sucesión de valores discretos preestablecidos según el código utilizado. (Faúndez Zanuy, 2000)

La cuantificación más comúnmente usada, es de 8 bits, mínimo requerido para una calidad baja, puede mejorarse su S/R con una técnica no lineal de cuantificación, se obtienen excelentes resultados aumentando la cuantificación a 16 bits. (Schroeder, 1999)

Ejemplo:

Utilizando 4 bits se pueden representar 16 valores diferentes.

Con 8 bits ya son 256 valores.

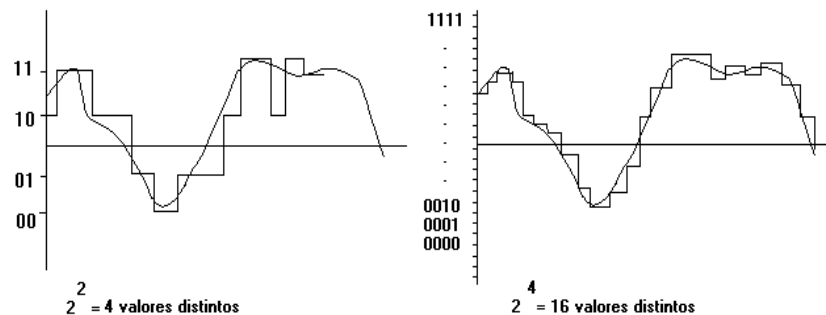


Figura 4. Señal Cuantizada con 2 y 4 bits.

Fuente. Faúndez Zanuy

Existen diferentes técnicas de cuantización:

- Cuantización uniforme
- Cuantización no uniforme
- Cuantización logarítmica
- Cuantización vectorial

1.5.1.7 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.

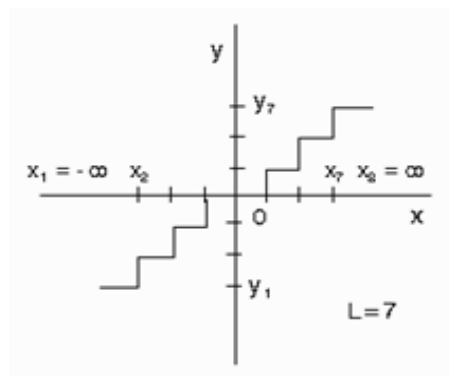


Figura 5. Señal Cuantizada Uniformemente

Fuente. Faundez Zanuy

1.5.1.8 CUANTIFICACIÓN NO UNIFORME

El problema de la cuantificación no uniforme es que conforme aumenta la amplitud de la señal, también aumenta el error. Este problema lo resuelve el cuantificador logarítmico de forma parcial. Sin embargo, si se conoce la función de la distribución de probabilidad, se puede ajustar los niveles de reconstrucción a la distribución de forma que se minimice el error cuadrático medio. Esto significa que la mayoría de los niveles de reconstrucción se den en las entradas más frecuentes y, consecuentemente, se minimice el error (ruido).

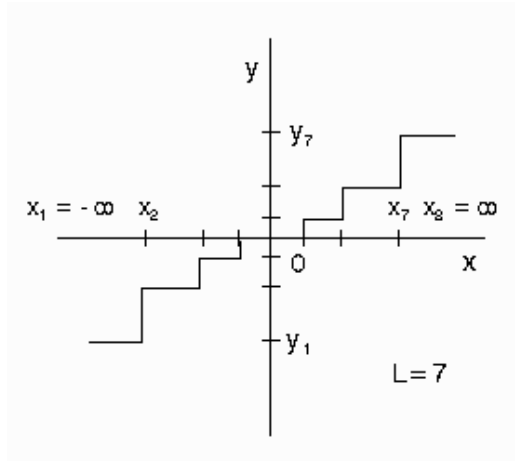


Figura 6. Señal Cuantizada No Uniformemente

Fuente. Faúndez Zanuy

1.5.1.9 CUANTIFICACIÓN LOGARÍTMICA

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

1.5.1.10 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. La cuantificación vectorial ofrece mejores resultados que la cuantificación escalar, sin embargo, es más sensible a los errores de transmisión.

1.5.1.11 MEDICIÓN DE CARACTERÍSTICAS.

Se trata básicamente de una técnica de reducción de datos en la cual el gran número de datos en la señal grabada es transformado en un grupo más pequeño de características que describen las propiedades importantes de la señal.

1.5.1.12 CONVERSIÓN A/D O ADQUISICIÓN

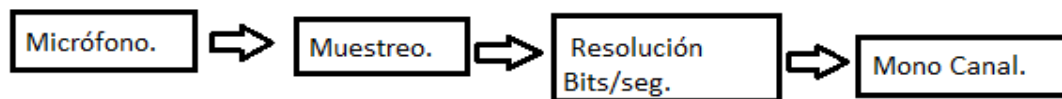


Figura 7. Diagrama de la adquisición de Voz

Fuente. Autores

1.5.1.13 OBTENCIÓN DE LA SEÑAL POR EL MICRÓFONO

El micrófono es un transductor electro-acústico. Su función es la de transformar la presión acústica ejercida sobre su capsula por las ondas sonoras en energía eléctrica. Cuando un micrófono está operando las ondas de sonido hacen que vibre el elemento magnético del micrófono causando una corriente eléctrica hacia la tarjeta de sonido, donde el convertidor A/D básicamente graba los voltajes eléctricos en intervalos específicos.

1.5.1.14 FILTRACIÓN

Consiste en quitar información o datos indeseados de entrada. Dependiendo del uso, el algoritmo o método de filtrado cambia.

1.5.1.15 ELIMINACIÓN DE RUIDO Y NORMALIZACIÓN

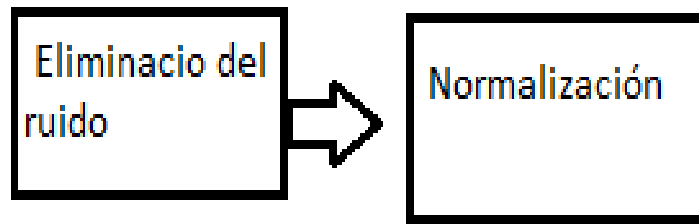


Figura 8. Diagrama de eliminación de ruido y Normalización

Fuente. Autores

Primero se realiza un filtrado de la señal para reducir los efectos del ruido que acompaña la señal de voz ingresada, sin este filtrado se tendrían datos que no son necesarios.

También se hace una segmentación, esto es detectar el inicio y final de la señal de voz para eliminar el ruido inicial y final, esto ayuda a optimizar memoria y hacer más veloz el tiempo de cálculo para comparaciones y por supuesto evitar errores, para hacer esto se hace uso del nivel de energía.

1.5.1.16 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.

1.6 COEFICIENTES CEPSTRALES DE FRECUENCIA DE MEL

MFCC está basado en el modelo del sistema auditivo periférico (Saha & Yadhunandan, 2004) en el cual la percepción de los contenidos de frecuencia no sigue un comportamiento lineal. Por esto se hace necesario realizar una medida del

contorno de la frecuencia fundamental sobre la escala Mel, que es una escala logarítmica para valores por encima de 1kHz y lineal por debajo de este valor. Para el cálculo de los MFCC fue empleada la función `mfcc.m` del *Auditorytoolbox* del software Matlab. (Aguilar Torres & Martínez Mascorro, 2012)

Los MFCCs representan elementos en el dominio de la frecuencia, que se utilizan para poder reducir en número a valores muy representativos. Los pasos a seguir son los siguientes:

- Dividir la señal en ventanas del mismo tamaño (se recomienda que estén intercaladas).
- De preferencia de una ventana de Hamming por cada ventana.
- Para cada ventana calcular la transformada discreta de Fourier.
- Con los valores de la transformada de Fourier calcular el periodo grama que representa la energía por cada valor de frecuencia.
- Filtrar las frecuencias y energías del periodo grama usando un banco de filtros Mel.
- Calcular el logaritmo de la energía por cada uno de los filtros.
- Calcular el transformado coseno discreto de las energías.
- Quedarse con los valores 1 a 13 de esta transformada, que representan a los 13 coeficientes MFCCs.

1.6.1 VENTANEO

Debido que nuestra señal de audio no es periódica no es posible calcular sus componentes en frecuencia para toda la señal. Por lo que se procede calcular los componentes para cada segmento. A estos segmentos se le conocen como ventanas. La ventana representa a una función periódica, por lo que se aplica fórmulas matemáticas que permite saber los componentes para una función periódica.

Una vez obtenidos los componentes para una ventana se puede mover a la siguiente ventana en el tiempo y calcular los componentes para esa nueva ventana.

1.6.2 HAMMING

La ventana Hamming ayuda a obtener la información en la ventana de la señal original, esto ayudará a filtrar frecuencias espurias que aparecen por cortar abruptamente la señal.

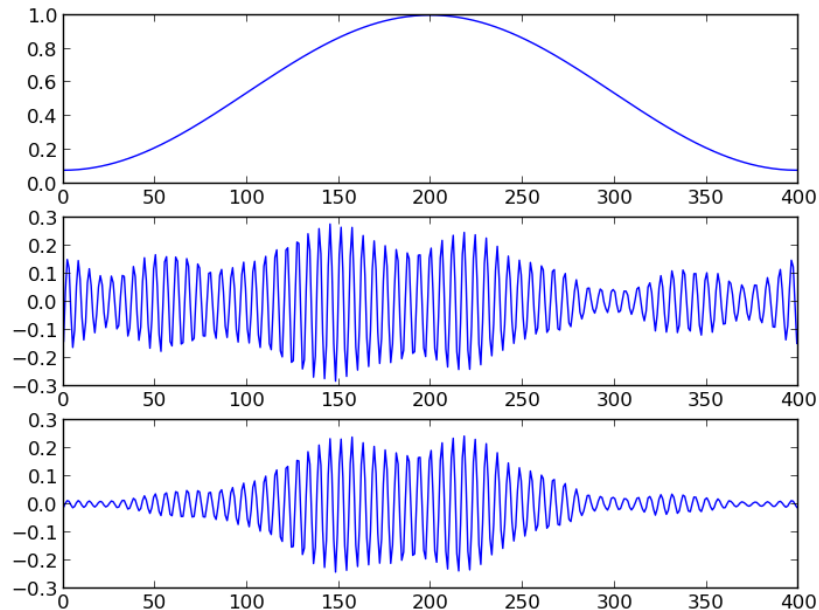


Figura 9. Ventana de Hamming

Fuente: Iván Vladimir Meza Ruiz

1.6.3 TRANSFORMADA DISCRETA DE FOURIER

La transformada discreta de Fourier es la operación que permite calcular los componentes en el dominio de la frecuencia. Desafortunadamente, la transformada da componentes en el dominio de los números imaginarios, que son los que se utilizan para representar información sobre la amplitud de las funciones periódicas por grupos de frecuencias. Sin embargo en lo interesa es en la energía en esas frecuencias, por lo que se eleva al cuadrado la información real e imaginaria para extraer la magnitud de la energía.

1.6.4 FILTROS MEL

Un filtro Mel es un filtro triangular que ayuda a obtener información de una banda de frecuencia. Estas bandas están basadas en la percepción del oído humano, en donde frecuencias bajas son más granulares y con mayor peso y frecuencias altas más amplias pero con menor peso.

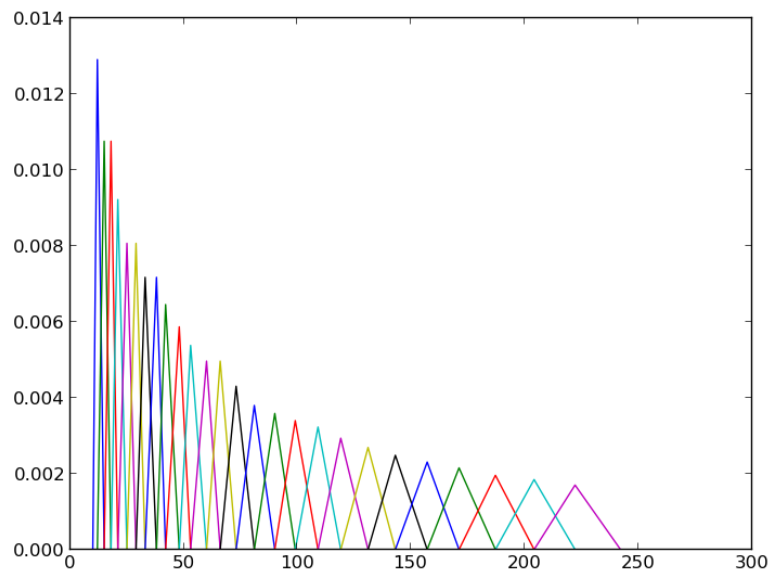


Figura 10. Banco de filtros

Fuente: Iván Vladimir Meza Ruiz

Como se puede apreciar un filtro MEL se concentra en las frecuencias bajas y abarca pocas frecuencias, conforme crece las frecuencias comienza abarcar más frecuencias pero la altura del filtro disminuye.

1.6.5 TRANSFORMADA COSENO DISCRETA

La escala logarítmica de la energía de los coeficientes MEL es la que convierte la señal en un coeficiente cepstral. El prefijo "ceps" es el inverso de "spec" (de espectral) y lo que señala es que tratar dicha señal como si fuera una señal en amplitud. Por eso, el siguiente paso es aplicar una transformada de coseno a los

coeficientes MEL, que permite comprimir dicha señal en elementos más informativos.

De esta forma es posible calcular los MFCCs para una señal de voz. El procedimiento se realizó de la siguiente manera: utilizando ventanas de Hamming, identificar la energía para los componentes en el dominio de la frecuencia, filtrar estos componentes con un modelo de la percepción humana y comprimir estos valores aún más. Los coeficientes resultantes pueden ser utilizados para tareas de aprendizaje automático en donde se les denominará como las características de la señal.

CAPÍTULO II

2 METODOLOGÍA

2.1 TIPO DE ESTUDIO

2.1.1 DESCRIPTIVO

La investigación experimental es un proceso que consiste en someter a un objeto o grupo de individuos a determinadas condiciones, estímulos o tratamiento (variable independiente), para observar los efectos o reacciones que se producen (variable dependiente). (Arias, 2006)

La investigación descriptiva consiste en la caracterización de un hecho, fenómeno, individuo o grupo, con el fin de establecer su estructura o comportamiento. Los resultados de este tipo de investigación se ubican en un nivel intermedio en cuanto a la profundidad de los conocimientos se refiere. (Arias, 2006)

El diseño es de un software se basa mediante una investigación de campo experimental ya que se realiza pruebas para comprobar el grado de factibilidad que tiene para reconocer los estados emocionales de las personas en la Universidad Nacional de Chimborazo, para analizar varios factores que pueda ayudar a mejorar su estilo de vida, este motivo se la ubica en una investigación descriptiva.

2.2 MÉTODOS, TÉCNICAS E INSTRUMENTOS

2.2.1 MÉTODOS

2.2.2 ANALÍTICO/DEDUCTIVO

Se utiliza el método analítico-deductivo en el proyecto ya que se inicia con un análisis particular del funcionamiento de los componentes necesarios para el

desarrollo del software y la forma de interactuar entre sí para que el sistema desempeñe su función de la manera óptima.

2.2.3 TÉCNICAS

2.2.3.1 OBSERVACIÓN

Se utilizó la técnica de la observación, que consiste en visualizar o captar mediante la vista, en forma sistemática, cualquier hecho, fenómeno o situación que se produzca en la naturaleza o en la sociedad, en función de los objetivos de investigación preestablecidos. Es estructurada porque se utiliza una guía diseñada previamente, en la que se especifican los elementos que serán observados. En este proyecto se observa los siguientes parámetros la frecuencia y tiempo.

2.2.3.2 INSTRUMENTACIÓN

Se utilizó listas de cotejo, folletos, libros, revistas y otros instrumentos necesarios para la selección de la información que se refiere al diseño e implementación del mismo.

2.3 POBLACIÓN MUESTRA

2.3.1 POBLACIÓN

La población es cualquier conjunto de elementos de los que se quiere conocer o investigar alguna o algunas de sus características. Se trabajó con una población de 20 estudiantes de la Universidad Nacional de Chimborazo de la Facultad Ingeniería de la carrera de Electrónica y Telecomunicaciones.

2.3.2 MUESTRA

Se trabajó con toda la población seleccionada el segundo semestre.

2.4 HIPÓTESIS

La elaboración de un software para detectar estados emocionales por medio de reconocimiento de tonos de voz y gramáticas permitirá una mejora en los parámetros interacción.

2.5 OPERACIONALIZACIÓN DE VARIABLES

Para la operacionalización de las variables se debe tomar en cuenta las variables independientes y las variables dependientes del tema descrito y la hipótesis planteada.

En la Tabla 2 se indica la operacionalización de variables dependientes e independientes del proyecto.

Tabla 2. Operacionalización de Variables

Fuente: Autores

VARIABLE	CATEGORÍA	INDICADOR	ÍTEM	TÉCNICA/INSTRUMENTO
<p><u>V. Independiente</u></p> <p>La elaboración de un software para detectar estados emocionales por medio de reconocimiento de tonos de voz y gramáticas.</p>	<p>Generalmente se ha trabajado con bases de datos actuadas las cuales proporcionan “Retratos de emociones” representando expresiones prototípicas e intensas que facilitan la búsqueda de correlación acústica y la subsecuente clasificación automática, sin embargo, no se han tenido buenos resultados al trasladar el conocimiento extraído de estas bases de datos a contextos reales. (Staild S., 2009)</p>	<p>Se analizara la entonación de voz de los estudiantes para de esta manera obtener información necesaria para la clasificación de los estados emocionales.</p>	<p>Identificar parámetros</p> <p>Identificar relaciones entre variables</p>	<p>Observación, guía de observación</p> <p>Encuestas</p> <p>Entrevista</p>
<p><u>V. Dependiente</u></p> <p>Mejora en los parámetros de interacción.</p>	<p>Se aplicarán diferentes métodos de interacción y de esta manera seleccionar los parámetros más efectivos para la interacción humano-maquina.</p>	<p>Se Medirá niveles y tiempos en los que se mantiene interacción tiempo real y de manera fluida.</p>	<p>Identificación de los componentes</p> <p>Identificar parámetros</p>	<p>Observación, guía de observación</p> <p>Encuestas</p> <p>Entrevista</p>

2.6 PROCEDIMIENTOS

Para el desarrollo del software de reconocimiento de estados emocionales, se siguió una serie de procesos, que, en conjunto, cumplirían con el objetivo deseado. A continuación, se presenta los siguientes pasos a seguir en la investigación propuesta:

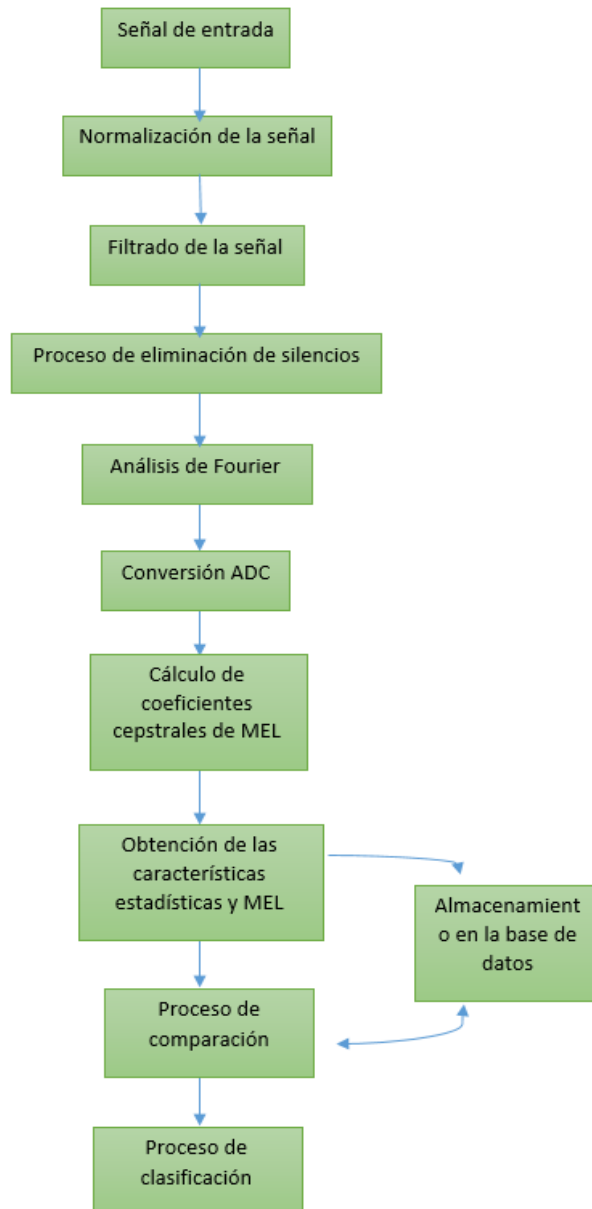


Figura 11. Funcionamiento del software

Fuente: Autores

2.7 PROCESAMIENTO

2.7.1 DISEÑO

La base de datos posee 5 grabaciones del usuario, las cuales se almacenan y en formato .WAV utilizando el software MATLAB a una frecuencia de muestreo de 44100 KHz y con 16 bits de resolución.

Posteriormente se diseñó el filtro para eliminar el ruido y aumentar la energía de la señal de audio para alcanzar los parámetros normalizados.

Se capturaron 5 señales de audio en el rango de frecuencia necesario para definir el estado de ánimo como Felicidad. El rango es (0-1037.96 Hz), con una frecuencia central media de 140.2350 Hz, y con una amplitud media de 84.4369 dB. De estas señales se obtuvo una señal promedio que será la señal “modelo” en base a lo que se definió es estado de ánimo de la persona.

Se capturaron 5 señales de audio en el rango de frecuencia necesario para definir el estado de ánimo como Tristeza. El rango es (0-570 Hz), con una frecuencia central media de 132.4292 Hz, y con una amplitud media de 83.6496 dB. De estas señales se obtuvo una señal promedio que será la señal “modelo” en base a lo que se definió es estado de ánimo de la persona.

Se capturaron 5 señales de audio en el rango de frecuencia necesario para definir el estado de ánimo como Miedo. El rango es (0- 511 Hz), con una frecuencia central media de 144.5416 Hz, y con una amplitud media de 84.9527 dB. De estas señales se obtuvo una señal promedio que será la señal “modelo” en base a lo que se definió es estado de ánimo de la persona.

Se capturaron 5 señales de audio en el rango de frecuencia necesario para definir el estado de ánimo como Sorpresa. El rango es (0- 585 Hz), con una frecuencia central media de 161.9028 Hz, y con una amplitud media de 82.6823 dB. De estas señales se obtuvo una señal promedio que será la señal “modelo” en base a lo que se definió es estado de ánimo de la persona.

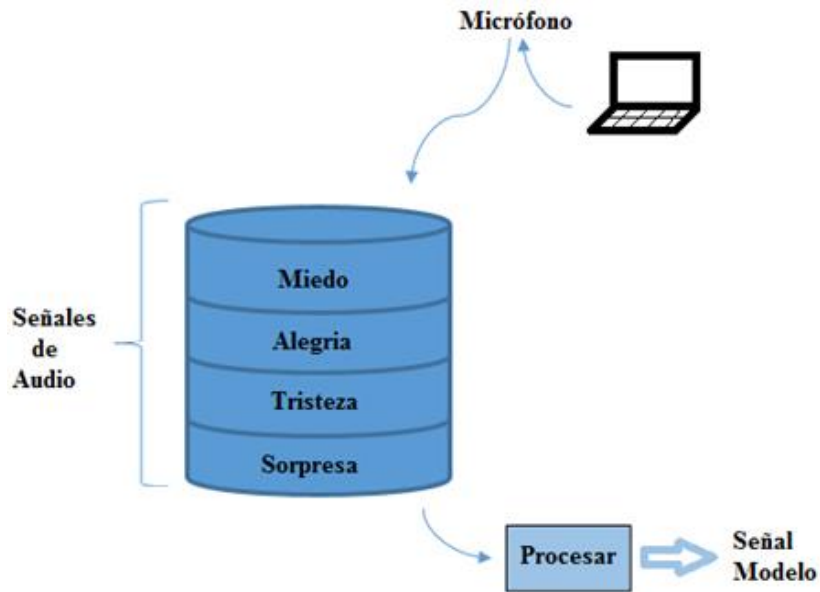


Figura 12. Procesamiento

Fuente. Autores

Se adquiere las 5 señales de audio y se obtiene la señal modelo, la cual debe ser normalizada.

2.8 FUNCIONES UTILIZADAS

Grabadora

Crea una entrada de audio

```
[y,fs]=grabadora;
```

Guardar

Almacena el audio grabado en formato wav y se le asigna un nombre

```
guardar(y, fs);
```

Cargar

Lee el audio grabado y su respectiva frecuencia de muestreo, almacenados con su respectivo nombre

```
[y, fs]=cargar(y);
```

Frecuencia

Utilizada para obtener el valor de la frecuencia

```
[frecuencia_Original, amplitud]=frecuencia_C(y, fs)
```

Longitud

Se obtiene la longitud del vector de la señal

```
L=longitud(y)
```

Normalizar

Normaliza la señal

```
[x3, ti]=normalizar(y, fs)
```

Silencio

Recorta los espacios en silencio y deja así solo la información útil

```
k=silencio(x3, fs)
```

Fourier

Utilizada para el análisis de la señal en función de la frecuencia

```
fourier(k, fs)
```

Características

Es utilizada para hallar las características de la señal

```
características(k, fs)
```

Muestreo

Se utiliza para muestrear la señal en la conversión ADC

```
[n_bits, Fs, senal_muestreada, index, a, muestras]=muestreo(k, fs);
```

Cuantización

Se utiliza para tomar valores de cuantización de acuerdo al número de bits seleccionados para el muestreo la señal en la conversión ADC

```
[flujo, tam, n_bits, niveles, a, muestras, vmin, vmax, fs, s_cuantizada]=cuantizacion(y, n_bits, index, senal_muestreada, a, muestras, fs, frecuencia_acotada, amplitud)
```

Codificación

Realiza la conversión de la señal cuantizada a su representación en codificación binaria.

```
[n_bits, Fs, senal_muestreada, index, a, muestras]=muestreo(k, fs);  
[flujo, tam, n_bits, niveles, a, muestras, vmin, vmax, fs, s_cuantizada]=cuantizacion(y, n_bits, index, senal_muestreada, a, muestras, fs, frecuencia_acotada, amplitud)  
codifica(flujo, tam, n_bits, niveles, a, muestras, vmin, vmax, fs, s_cuantizada, frecuencia_acotada, amplitud);
```

Base de datos

Realiza el proceso de almacenamiento en una base de datos.

```
Base_de_datos;
```

Proceso

Realiza el proceso de clasificación de estados emocionales.

```
Proceso(y, fs)
```

2.8.1.1.1 DESARROLLO

Al iniciar el programa aparecerán 3 opciones, la de crear una base de datos para el reconocimiento de los estados emocionales, la de reconocimiento y clasificación, y la de salir del programa.

Al seleccionar la opción de crear la base de datos se siguen los siguientes pasos:

- Grabación
- Guardar
- Cargar
- Frecuencia
- Longitud
- Normalizar
- Silencio
- Fourier
- Características estadísticas
- Características de Mel

En la opción de clasificar un estado emocional los pasos a seguir son los siguientes

- Grabación
- Guardar
- Cargar
- Frecuencia
- Longitud
- Normalizar
- Silencio
- Fourier
- Características estadísticas
- Características de Mel
- Comparación
- Proceso

En el último paso se realiza el proceso de clasificación de los estados emocionales de las personas

```
clc;clear all;
disp ('=====');
disp ('=          Escoja una opción          =');
disp ('=====');
opcion=input('1-Crear Base de Datos \n2-Clasificar
Estados (Hombre) \n3-Clasificar Estados (Mujer) \n4-
Salir\n');
clc;
switch opcion
%%
    case 1,
        [y,fs]=grabadora;
        nom=guardar(y,fs);
        [y,fs]=cargar(y,nom);
        L=longitud(y);
        [x3,ti]=normalizar(y,fs);
        [senal_filtrada,filtb,filta]=pasabajos_butterworth(x3,fs);
        k=silencio(senal_filtrada,fs,nom);
        fourier(k,fs,nom);
        [frecuencia_acotada,amplitud]=frecuencia_C(k,fs);
        [fs,mediana,Longitud,media,maximo,minimo,desviacion_estandar,
        Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,amplitud]=
        caracteristicas(k,fs);
        muestreo(k,fs,frecuencia_acotada,amplitud);
        Base_de_datos;
        clc;
        %%
    case 2 ,
        [y,fs]=grabadora;
```

```

nom=guardar(y, fs);
[y, fs]=cargar(y, nom);
L=longitud(y);
[x3, ti]=normalizar(y, fs);
[senal_filtrada, filtb, filta]=pasabajos_butterworth(x3, fs);
k=silencio(senal_filtrada, fs, nom);
fourier(k, fs, nom);
[frecuencia_acotada, amplitud]=frecuencia_C(k, fs);
[ MFCCs, FBEs, frames ]=mel(k, fs);
[fs, mediana, Longitud, media, maximo, minimo, desviacion_estandar, Varianza, Curtosis, Asimetria, Vp, Sh, Cap, frecuencia, amplitud]=caracteristicas(k, fs);
[medianamfcc, Longitudmfcc, mediamfcc, maximomfcc, minimomfcc, desviacion_estandarmfcc, Varianzmfcc, Curtosismfcc, Asimetriamfcc, Vpmfcc, Shmfcc, Capmfcc]=caracteristicas_Mel(MFCCs);
muestreo(k, fs, frecuencia_acotada, amplitud);
process;
%%
    case 3 ,
[y, fs]=grabadora;
nom=guardar(y, fs);
[y, fs]=cargar(y, nom);
L=longitud(y);
[x3, ti]=normalizar(y, fs);
[senal_filtrada, filtb, filta]=pasabajos_butterworth(x3, fs);
k=silencio(senal_filtrada, fs, nom);
fourier(k, fs, nom);
[frecuencia_acotada, amplitud]=frecuencia_C(k, fs);
[ MFCCs, FBEs, frames ]=mel(k, fs);

```



```

[fs,mediana,Longitud,media,maximo,minimo,desviacion_est
andar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,
amplitud]=caracteristicas(k,fs);
muestreo(k,fs,frecuencia_acotada,amplitud);
process;
%%
    case 4,
disp('Cerrando el programa')
close all;
pause(2);
clc
fuera=1;
end

```

2.8.2 ENTRENAMIENTO Y BASE DE DATOS

2.8.2.1 CAPTURA DEL AUDIO

Primero se crea un objeto de audio con $y = \text{wavrecord}(t * fs, fs, 1)$; donde “t” es igual al tiempo en segundos para la duración de la grabación del audio, “fs” es la frecuencia a la que se realizará el muestreo de la señal, en nuestro caso se utiliza una fs de 44100 Hz y el “1” indica que se utilizara un solo canal o mono y con un valor de 16 bits/muestra por defecto.

Luego de la grabación del audio se procede a guardarlo, el programa solicitará un nombre para almacenar el audio, el cual deberá ser escrito por el usuario $\text{wavwrite}(y, fs, 16, \text{nom})$; donde “y” es el vector con la información de la señal que se va a guardar, “fs” es la frecuencia de muestreo a la que se guardara la señal, 16 es el número de bit por muestra a los que se guardara la señal, y “nom” es la variable tipo string, que almacenará el nombre para el archivo de audio.

Luego se procede a cargar el audio almacenado para adquirir su información necesaria almacenada como un vector de datos, su frecuencia de muestreo, y su número de bits.

En esta etapa se visualiza la señal de la voz en función del tiempo, ya que se realizó una grabación de 2 segundos, la señal se visualiza en este periodo de tiempo.

```
function [y,fs]=grabadora
clc;clear all;
pause;
fs=44100;
t=2;
y=wavrecord(t*fs,fs,1);
sound(y,fs);
disp('Grabación realizada');
```

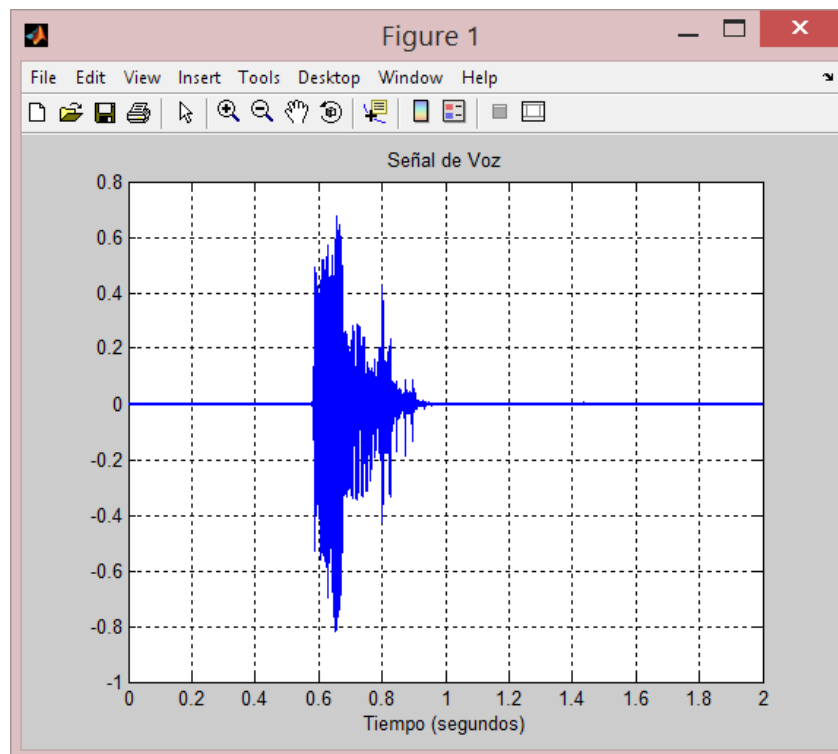


Figura 13. Audio Capturado

Fuente. Autores

2.8.2.2 NORMALIZACIÓN

En esta etapa se realiza la normalización de la señal porque todo software requiere que la señal sea normalizada para su posterior análisis.

```
function [x3,ti]=normalizar(y,fs)
```

```

%% NORMALIZACION DE LA SEÑAL "x2 es la señal
normalizada"
L=length(y);
Ts=1/fs; % Tiempo de muestreo.
ti=(1:L)*Ts; % Vector de tiempo.
x=y;
mwad=max(x);
mwid=min(x);
x3=x/(max(mwad,mwid));
plot (ti,x3)
sound(x3,fs)
guardar(x3,fs)
disp('Normalización realizada');

```

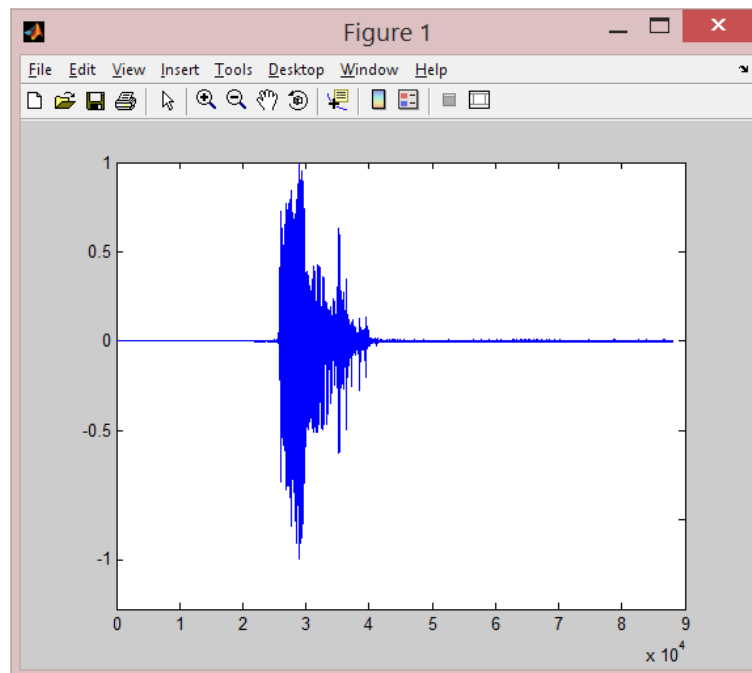


Figura 14. Normalización

Fuente. Autores

2.8.2.3 ELIMINACIÓN DE SILENCIOS

En esta etapa se aplica un algoritmo para eliminar el silencio de nuestra señal para tener únicamente la parte de la señal con la información de voz útil para el procesamiento.

```
function k=silencio(y2,fs)
disp('precione ENTER para continuar con la eliminacion
del silencio');
pause;
wavplay(y2,fs);
s=y2;
figure(2)
subplot(211)
plot(y2)
len = length(s); % longitud del vector
d=max(abs(s));
s=s/d;
avg_e = sum(s.*s)/len;
THRES = 0.02;
y2 = [0];
for i = 1:400:len-400
seg = s(i:i+399);
e = sum(seg.*seg)/400;
if( e> THRES*avg_e)
y2=[y2;seg(1:end)];
end;
end;
figure(2)
subplot(212)
plot(y2)
sound(y2,fs)
k=y2;
nom=input('escriba un nombre para el audio: ','s');
```

```
wavwrite(y2, fs, 16, nom);
display('el audio se ha guardado correctamente');
```

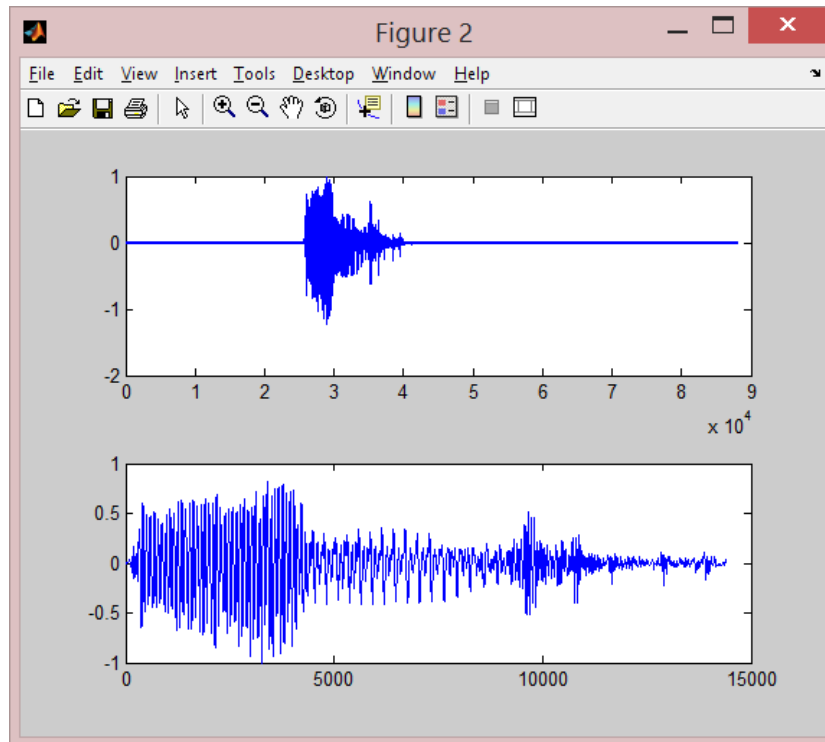


Figura 15. Eliminación de silencios

Fuente. Autores

2.8.2.4 FRECUENCIA

En esta etapa se realiza el cálculo para pasar la señal del tiempo a la frecuencia utilizando la transformada rápida de Fourier.

```
function [freq, amp]=frecuencia_C(y, Fs);
f=linspace(-Fs/2, Fs/2, length(y)); %Vector creado para
el eje de frecuencias
sa=fftshift(abs(fft(y))); %transformada rápida de
fourier a la cual se le saca el valor absoluto
```

```

fila1=sa.' % Organiza los datos en vector fila.
amp=max(sa); % Pico máximo
subplot(211)
plot(f,fftshift(abs(fft(y))))%reorganiza las salidas de
FFT, moviendo el componente de frecuencia cero hasta el
centro de la matriz.
%Es útil para la visualización de una transformada de
Fourier con el componente de frecuencia cero en el
centro del espectro.
w=find(sa==amp);%encuentra la posición de
mw=max(w); %halla el valor más alto
freq=f(mw)%halla la frecuencia central haciendo una
relación con la amplitud pico del espectro de
frecuencia
disp('Cálculo de frecuencia realizado');
%%
fs=Fs;
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiete potencia de 2 de la
long de y
Y = fft(y,NFFT)/L; % FFT de la señal
f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1))./4e-10);
amplitud=max(FDB)
xlim([0,5000])
subplot(212)
plot(f,FDB)%%%%%%%%%%%
xlim([0,5000])
title('Espectro unilateral de y(t)')
xlabel('Frecuencia (Hz)')
ylabel('dB')
w=find(FDB==amplitud);%encuentra la posición de
mw=max(w); %halla el valor más alto

```

```
frecuencia=f(mw)%halla la frecuencia central haciendo  
una relación con la amplitud pico del espectro de  
frecuencia  
pause;
```

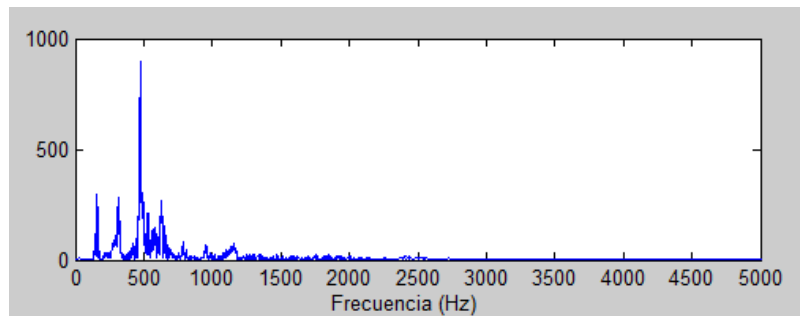


Figura 16. Espectro de frecuencia

Fuente. Autores

También se realiza el análisis de la señal en dBs, y de esta manera obtener sus valores de amplitud pico en dB, y de su frecuencia central.

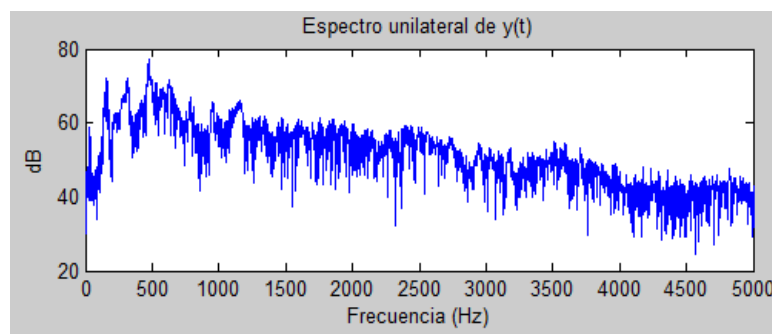


Figura 17. Espectro de frecuencia en dBs

Fuente. Autores

2.8.2.5 MUESTREO

Se realiza la conversión analogía a digital, para lo cual se realiza el muestreo, el cual consiste en tomar muestras de la señal de entrada para su posterior análisis con la cuantización.

```

n_bits1=input('ingrese el número de bits: ','s');
Fs1=input('ingrese el número de muestras: ','s');
n_bits=str2double(n_bits1);
Fs=str2double(Fs1);
disp('Procesando...');
[a,il]=size(y);
tiempo=a/fs;
Fs=round(Fs*tiempo);
figure(3)
plot(y);
wavplay(y,fs);
%% -----Muestreo-----
muestras=round(linspace(1,a,Fs));
senal_muestreada=zeros(1,Fs);
index=1;
for i=muestras
    senal_muestreada(index)=y(i);
    index=index+1;
end
stem(senal_muestreada);
disp('Muestreo realizado');

```

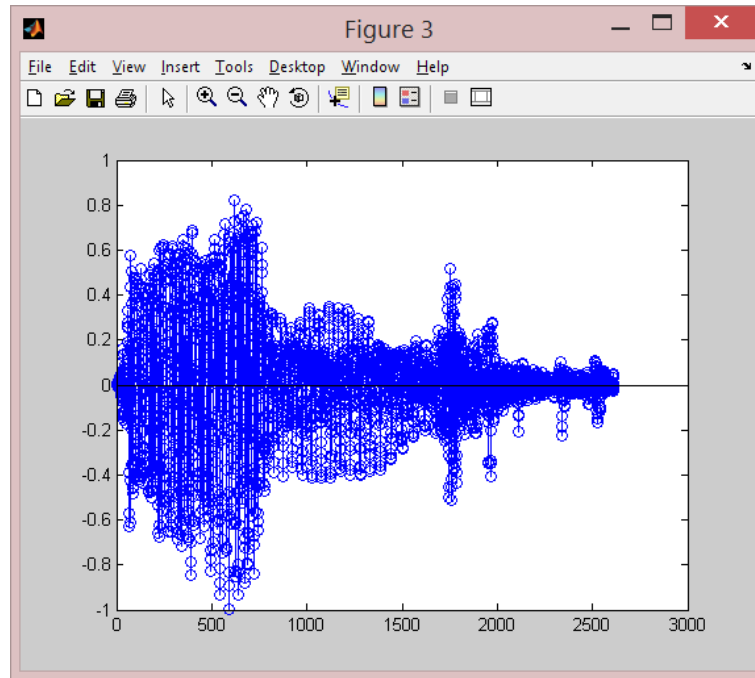



Figura 18. Señal muestreada

Fuente. Autores

2.8.2.6 CUANTIZACIÓN

Se procede a realizar la cuantización de la señal de entrada.

```
function
cuan=cuatizacion(y,n_bits,index,senal_muestreada,a,muestras,fs,frecuencia_acotada,amplitud)
%% -----Cuantización-----
vmax=max(y);
vmin=min(y);
lsb=(vmax-vmin)/(2^n_bits);
niveles=vmin:lsb:vmax;
particion=(vmin-(0.5*lsb)):lsb:(vmax+(0.5*lsb));
[il,indice_niv]=size(niveles);
% genera valores cuantizados de muestras
s_cuantizada=zeros(1,index-1);
%=====
for i=1:index-1
```

```

    for j=1:indice_niv
% si el valor es <vmin entonces sera igual a vmin
        if (senal_muestreada(i)<particion(1))
            s_cuantizada(i)=vmin;
            val_binairo=dec2bin(0,n_bits);

            rango_val=val_binario(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end
        if (senal_muestreada(i)>particion(end))
            s_cuantizada(i)=vmax;
            val_binairo=dec2bin(2^n_bits)-1,n_bits;

            rango_val=val_binario(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end

        if (senal_muestreada(i)>=particion(j))
            if (senal_muestreada(i)<particion(j+1))
                s_cuantizada(i)=niveles(j);
                %Al mismo tiempo la generaci3n de
flujo binario para cada
                %valor cuantificado
                val_binairo=dec2bin(j-1,n_bits);

                rango_val=val_binairo(:);
                rango_index=((i-1)*n_bits)+1;
                flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;

```

```

        end
    end
end
end
[tam,il]=size(flujo);
figure(4)
stairs(s_cuantizada)
ylim([vmin vmax])
set(gca,'YTick',vmin:lsb:vmax)
disp('Cuantización realizada');

```

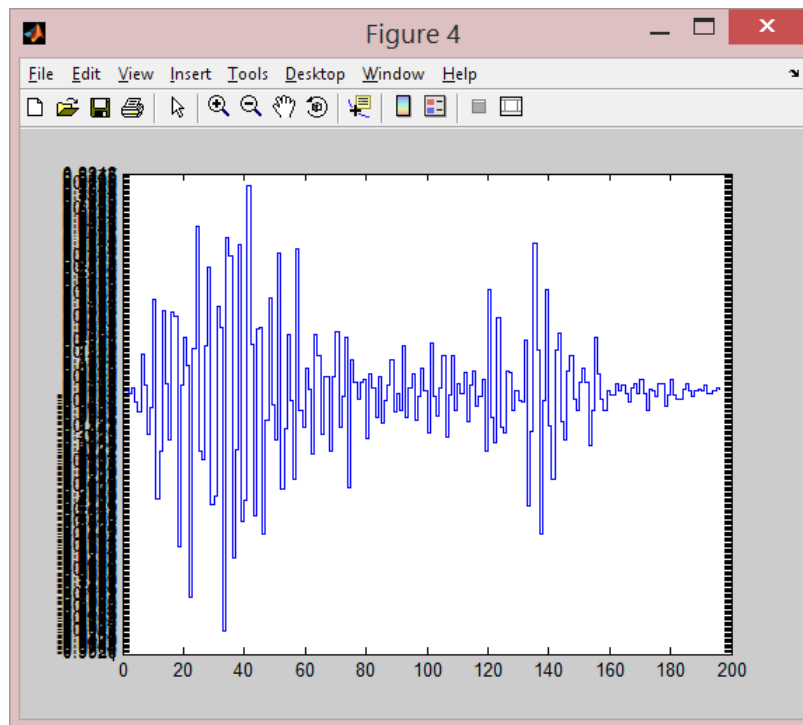


Figura 19. Señal cuantizada

Fuente. Autores

2.8.2.7 CODIFICACION

En esta etapa se realiza la conversión de la señal cuantizada a su representación binaria

```

val_cod=zeros(tam,1);
for i=1:tam
    val_cod(i,1)=str2double(flujo(i,1));
end
figure(2)
subplot(212)
stairs(val_cod');
axis([1 1000 -3 4]);
set(gca,'YTick',-3:1:4);
set(gca,'YTicklabel',{' ',' ',' ','0','1',' ',' ',' ',' '});
');

```

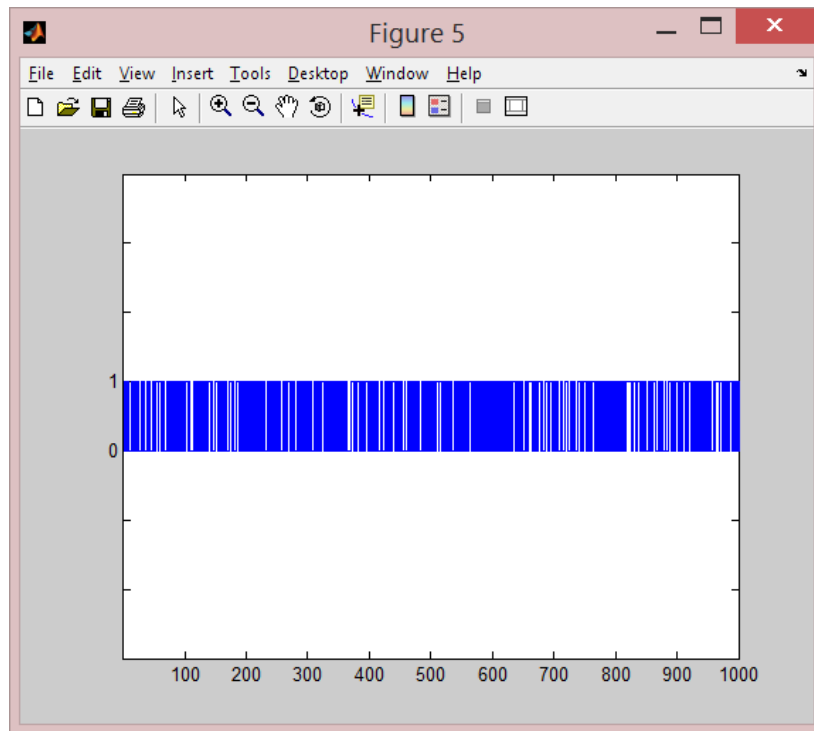


Figura 20. Señal codificada

Fuente. Autores

2.8.2.8 OBTENCION DE CARACTERISTICAS ESTADISTICAS

Se calcularon características de la señal para realizar el análisis poder almacenar en la base de datos las mismas y así ayudar en el proceso de clasificación de los estados emocionales.

```
function
[fs,mediana,Longitud,media,maximo,minimo,desviacion_est
andar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,
amplitud]=caracteristicas(y,fs)
%[y,fs]=wavread('USUARIO');
disp('Caracteristicas de la voz');
pause;
sound(y,fs)
Fs=fs;
%% parametros
mediana=median(y)%calcula la mediana
Longitud=length(y); %Longitud de y
media=mean(y)%calcula la media
maximo=max(y)%calcula el valor mas alto
minimo=min(y)%calcula el valor mas bajo
desviacion_estandar=std(y)%calcula la desviacion
estandar
Varianza = var(y)
Curtosis = kurtosis(y) % calculo de la curtosis
Asimetria=skewness(y) % calculo la asimetria
%http://probestunalmz1.wikispaces.com/cap\_1\_p\_11
% y= numero total de muestras del vector
for i=1:(length(y)-1)
    var1=abs(y(i+1)-y(i));
end
V1=(1/((length(y)-1))*var1;
for i=1:(length(y)-1)
    var2=abs(y(i));
```

```

end
V2=(1/((length(y))-1))*var2;
Vp=V1/V2; %perturbacion de la amplitud

for j=1:(length(y)-1)
    val=abs(max(y(j+1))-max(y(j)));
end
V3=(1/((length(y))-1))*val;
for j=1:(length(y)-1)
    va2=max(abs(y(j)));
end
V4=(1/((length(y))-1))*va2;
Sh=V3/V4 %cálculo de Shimer
%% -----
for k=1:(length(y)-1)
    v1=(y(k+1)-y(k))^2;
end
V5=sqrt((1/((length(y))))*v1);
for i=1:(length(y)-1)
    v2=abs(y(k));
end
V6=(1/((length(y))-1))*v2;
Cap=V5/V6 %perturbacion de la amplitud
f=linspace(-Fs/2,Fs/2,length(y)); %Vector creado para
el eje de frecuencias
sa=fftshift(abs(fft(y))); %transformada rápida de
Fourier a la cual se le saca el valor absoluto
fila1=sa.' % Organiza los datos en vector fila.
amp=max(sa); % Pico maximo
subplot(211)
plot(f,fftshift(abs(fft(y))))%reorganiza las salidas de
FFT, moviendo el componente de frecuencia cero hasta el
centro de la matriz.

```

```

%Es útil para la visualización de una transformada de
Fourier con el componente de frecuencia cero en el
centro del espectro.
w=find(sa==amp);%encuentra la posición de
mw=max(w); %halla el valor más alto
freq=f(mw)%halla la frecuencia central haciendo una
relación con la amplitud pico del espectro de
frecuencia
disp('Cálculo de frecuencia realizado');
%%
fs=Fs;
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiente potencia de 2 de la
long de y
Y = fft(y,NFFT)/L; % FFT de la señal
f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1))./4e-10);
amplitud=max(FDB);
w=find(FDB==amplitud);%encuentra la posición de
mw=max(w); %halla el valor más alto
frecuencia=f(mw)%halla la frecuencia central haciendo
una relación con la amplitud pico del espectro de
frecuencia

nam=input('Ingrese un nombre para almacenar las
características: ','s');
save
(nam,'fs','mediana','Longitud','media','maximo','minimo
','desviacion_estandar','Varianza','Curtosis','Asimetri
a','Vp','Sh','Cap','frecuencia','amplitud');
disp('Parámetros estadísticos obtenidos correctamente')
pause;

```

2.8.2.9 OBTENCION DE CARACTERISTICAS DE MEL

En esta etapa se obtiene las características de los MFCCs para los coeficientes cepstrales calculados, y así poder almacenarlos en la base de datos generada en MATLAB.

```
function
[medianamfcc,Longitudmfcc,mediamfcc,maximomfcc,minimomf
cc,desviacion_estandarmfcc,Varianzmfcc,Curtosismfcc,As
imetriamfcc,Vpmfcc,Shmfcc,Capmfcc]=caracteristicas_Mel(
MFCCs)
clc;
disp('Caracteristicas de los parametros MEL');
pause;
%% parametros
medianamfcc=median(MFCCs);%calcula la mediana
Longitudmfcc=length(MFCCs); %Longitud de MFCCs
mediamfcc=mean(MFCCs);%calcula la media
maximomfcc=max(MFCCs);%calcula el valor mas alto
minimomfcc=min(MFCCs);%calcula el valor mas bajo
desviacion_estandarmfcc=std(MFCCs);%calcula la
desviacion estandar
Varianzmfcc = var(MFCCs);
Curtosismfcc = kurtosis(MFCCs); % cálculo de curtosis
Asimetriamfcc=skewness(MFCCs); % cálculo la asimetria
%http://probestunalmzl.wikispaces.com/cap\_1\_p\_11
% MFCCs= número total de muestras del vector
for i=1:(length(MFCCs)-1)
    var1=abs(MFCCs(i+1)-MFCCs(i));
end
V1=(1/((length(MFCCs))-1))*var1;
for i=1:(length(MFCCs)-1)
    var2=abs(MFCCs(i));
end
```



```

V2=(1/((length(MFCCs))-1))*var2;
Vpmfcc=V1/V2; %perturbacion de la amplitud

for j=1:(length(MFCCs)-1)
    va1=abs(max(MFCCs(j+1))-max(MFCCs(j)));
end
V3=(1/((length(MFCCs))-1))*va1;
for j=1:(length(MFCCs)-1)
    va2=max(abs(MFCCs(j)));
end
V4=(1/((length(MFCCs))-1))*va2;
Shmfcc=V3/V4; %cálculo de Shimer

for k=1:(length(MFCCs)-1)
    v1=(MFCCs(k+1)-MFCCs(k))^2;
end
V5=sqrt((1/((length(MFCCs))))*v1);
for i=1:(length(MFCCs)-1)
    v2=abs(MFCCs(k));
end
V6=(1/((length(MFCCs))-1))*v2;
Capmfcc=V5/V6 %perturbación de la amplitud
%%
nam=input('Ingrese un nombre para almacenar las
características: ','s');
save
(nam,'medianamfcc','Longitudmfcc','mediamfcc','maximomf
cc','minimomfcc','desviacion_estandarmfcc','Varianzamfc
c','Curtosismfcc','Asimetriamfcc','Vpmfcc','Shmfcc','Ca
pmfcc');
disp('Parámetros estadísticos obtenidos correctamente')
pause;

```

2.8.2.10 PROMEDIOS

En esta etapa se organizan todas las características de cada emoción y se calcula la media en un vector para luego realizar el reconocimiento de las emociones.

```
clc;
nom_1=input('Escriba el nombre de la primera
muestra:', 's');
[rec1,fs]=wavread(nom_1);
nom_2=input('Escriba el nombre de la segunda
muestra:', 's');
[rec2,fs]=wavread(nom_2);
nom_3=input('Escriba el nombre de la tercera
muestra:', 's');
[rec3,fs]=wavread(nom_3);
nom_4=input('Escriba el nombre de la cuarta
muestra:', 's');
[rec4,fs]=wavread(nom_4);
nom_5=input('Escriba el nombre de la quinta
muestra:', 's');
[rec5,fs]=wavread(nom_5);
% Crea vector final de grabación
prom=rec1;
tam=length(rec5);
for i=1:1:tam

prom(i)=(rec1(i)+rec2(i)+rec3(i)+rec4(i)+rec5(i))/5;
end
lon=length(prom);
d=max(abs(prom));
prom=prom/d;
%%Guarda .wav señal promediada
```

```

nom_prom=input('Escriba el nombre de la señal
promediada:', 's');
wavwrite(prom, fs, 16, nom_prom);
plot(prom)
sound(prom, fs)

```

2.8.2.11 BASE DE DATOS

Se procede a crear la base de datos para cada emoción, asignándole un nombre en un archivo “.mat”, almacenando dentro del mismo las características mencionadas pero de su valor promedio para cada emoción.

```

('feliz1.mat', 'fs', 'mediana', 'Longitud', 'media', 'maximo',
'minimo', 'desviacion_estandar', 'Varianza', 'Curtosis',
'Asimetria', 'Vp', 'Sh', 'Cap', 'frecuencia', 'amplitud')
Feliz1=[fs, mediana, Longitud, media, maximo, minimo, desviac
ion_estandar, Varianza, Curtosis, Asimetria, Vp, Sh, Cap, frec
uencia, amplitud];
load
('feliz2.mat', 'fs', 'mediana', 'Longitud', 'media', 'maximo',
'minimo', 'desviacion_estandar', 'Varianza', 'Curtosis',
'Asimetria', 'Vp', 'Sh', 'Cap', 'frecuencia', 'amplitud')
Feliz2=[fs, mediana, Longitud, media, maximo, minimo, desviac
ion_estandar, Varianza, Curtosis, Asimetria, Vp, Sh, Cap, frec
uencia, amplitud];
load
('feliz3.mat', 'fs', 'mediana', 'Longitud', 'media', 'maximo',
'minimo', 'desviacion_estandar', 'Varianza', 'Curtosis',
'Asimetria', 'Vp', 'Sh', 'Cap', 'frecuencia', 'amplitud')
Feliz3=[fs, mediana, Longitud, media, maximo, minimo, desviac
ion_estandar, Varianza, Curtosis, Asimetria, Vp, Sh, Cap, frec
uencia, amplitud];

```

```

load
('feliz4.mat','fs','mediana','Longitud','media','maximo',
', 'minimo','desviacion_estandar','Varianza','Curtosis',
'Asimetria','Vp','Sh','Cap','frecuencia','amplitud')
Feliz4=[fs,mediana,Longitud,media,maximo,minimo,desviac
ion_estandar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frec
uencia,amplitud];
load
('feliz5.mat','fs','mediana','Longitud','media','maximo',
', 'minimo','desviacion_estandar','Varianza','Curtosis',
'Asimetria','Vp','Sh','Cap','frecuencia','amplitud')

Feliz5=[fs,mediana,Longitud,media,maximo,minimo,desviac
ion_estandar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frec
uencia,amplitud];
Feliz=[Feliz1;Feliz2;Feliz3;Feliz4;Feliz5];
FelizT=mean(Feliz);

fs=FelizT(1);mediana=FelizT(2);Longitud=FelizT(3);media
=FelizT(4);maximo=FelizT(5);minimo=FelizT(6);
desviacion_estandar=FelizT(7);Varianza=FelizT(8);Curtos
is=FelizT(9);Asimetria=FelizT(10);Vp=FelizT(11);
Sh=FelizT(12);Cap=FelizT(13);frecuencia=FelizT(14);ampl
itud=FelizT(15);

save('Target_1','fs','mediana','Longitud','media','maxi
mo','minimo','desviacion_estandar','Varianza','Curtosis
','Asimetria','Vp','Sh','Cap','frecuencia','amplitud')

```

2.8.3 ADQUISICIÓN

2.8.3.1 CAPTURA DEL AUDIO

Se realiza la captura del audio a una fs de 44100 Hz de 2 segundos, con 16bits por muestra

```
function [y,fs]=grabadora
clc;clear all;
pause;
fs=44100;
t=2;
y=wavrecord(t*fs,fs,1);
sound(y,fs);
disp('Grabación realizada');
```

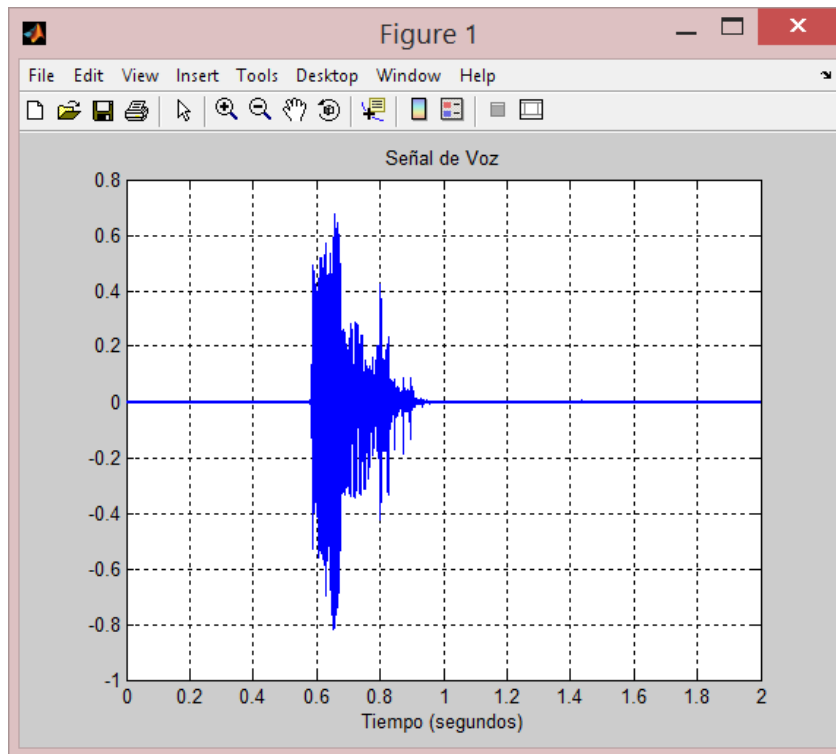


Figura 21. Audio Capturado

Fuente. Autores

2.8.3.2 NORMALIZACIÓN

Se realiza la normalización de la señal ajustando su valor máximo y normalizándolo a 1.

```
function [x3,ti]=normalizar(y,fs)
%% NORMALIZACION DE LA SEÑAL "x2 es la señal
normalizada"
L=length(y);
Ts=1/fs; % Tiempo de muestreo.
ti=(1:L)*Ts; % Vector de tiempo.
x=y;
mwad=max(x);
mwid=min(x);
x3=x/(max(mwad,mwid));
plot (ti,x3)
sound(x3,fs)
guardar(x3,fs)
disp('Normalización realizada');
```

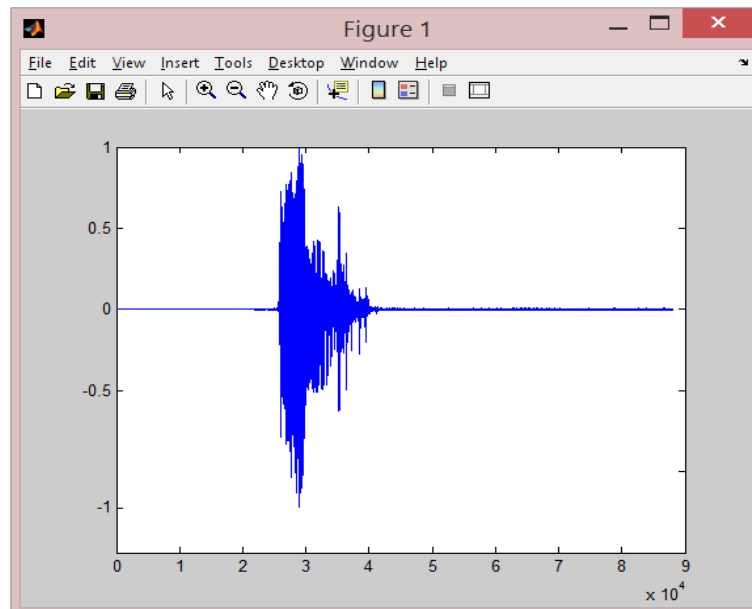


Figura 22. Normalización

Fuente. Autores

2.8.3.3 ELIMINACIÓN DE SILENCIOS

Por medio de la segmentación y ventanas de hamming se elimina el ruido y los silencios para que solo se tenga la señal pura que contenga la información de la voz.

```
function k=silencio(y2,fs)
disp('precione ENTER para continuar con la eliminacion
del silencio');
pause;
wavplay(y2,fs);
s=y2;
figure(2)
subplot(211)
plot (y2)
len = length(s); % longitud del vector
d=max(abs(s));
s=s/d;
avg_e = sum(s.*s)/len;
THRES = 0.02;
y2 = [0];
for i = 1:400:len-400
seg = s(i:i+399);
e = sum(seg.*seg)/400;
if( e> THRES*avg_e)
y2=[y2;seg(1:end)];
end;
end;
figure(2)
subplot(212)
plot(y2)
sound(y2,fs)
k=y2;
```

```

nom=input('escriba un nombre para el audio: ','s');
wavwrite(y2,fs,16,nom);
display('el audio se ha guardado correctamente');

```

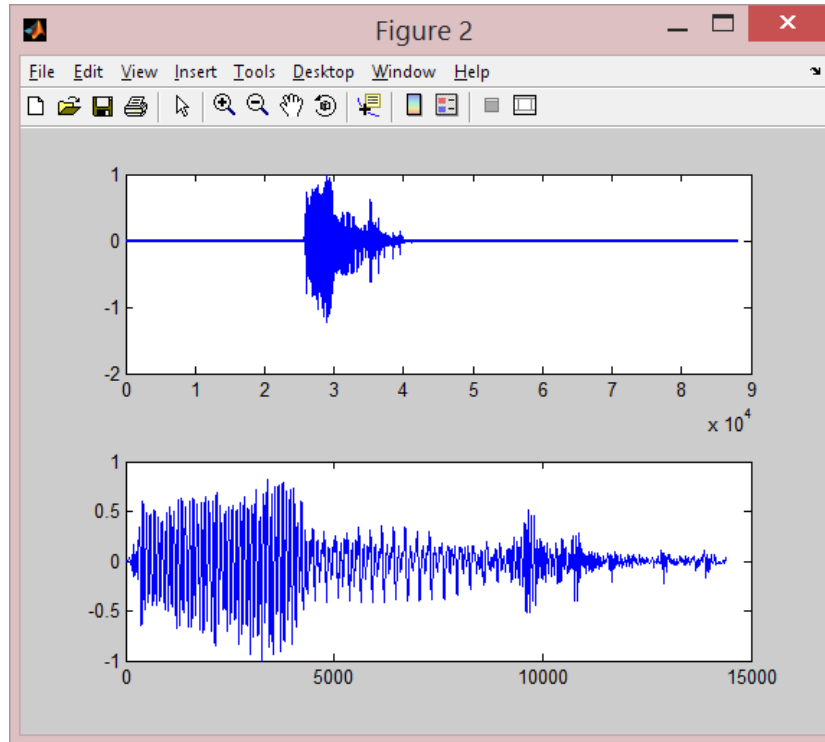


Figura 23. Eliminación de silencios

Fuente. Autores

2.8.3.4 FRECUENCIA

Se utiliza la transformada rápida de Fourier para el cálculo de la frecuencia.

```

function [freq, amp]=frecuencia_C(y, Fs);
f=linspace(-Fs/2,Fs/2,length(y)); %Vector creado para
el eje de frecuencias
sa=fftshift(abs(fft(y))); %transformada rapida de
fourier a la cual se le saca el valor absoluto
fila1=sa.' % Organiza los datos en vector fila.
amp=max(sa); % Pico maximo
subplot(211)

```



```

plot(f,fftshift(abs(fft(y))))%reorganiza las salidas de
FFT, moviendo el componente de frecuencia cero hasta el
centro de la matriz.
%Es útil para la visualización de una transformada de
Fourier con el componente de frecuencia cero en el
centro del espectro.
w=find(sa==amp);%encuentra la posicion de
mw=max(w); %halla el valor mas alto
freq=f(mw)%halla la frecuencia central haciendo una
relacion con la amplitud pico del espectro de
frecuencia
disp('Cálculo de frecuencia realizado');
%%
fs=Fs;
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiete potencia de 2 de la
long de y
Y = fft(y,NFFT)/L; % FFT de la señal
f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1)) ./4e-10);
amplitud=max(FDB)
xlim([0,5000])
subplot(212)
plot(f,FDB)%%%%%%%%%%%
xlim([0,5000])
title('Espectro unilateral de y(t)')
xlabel('Frecuencia (Hz)')
ylabel('dB')
w=find(FDB==amplitud);%encuentra la posicion de
mw=max(w); %halla el valor mas alto
frecuencia=f(mw)%halla la frecuencia central haciendo
una relacion con la amplitud pico del espectro de
frecuencia

```

```
pause;
```

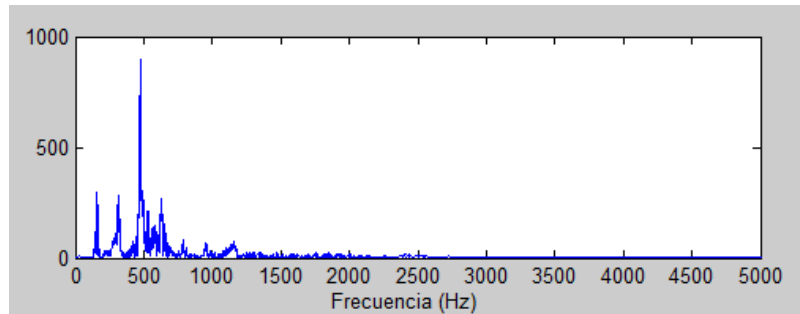


Figura 24. Espectro de frecuencia

Fuente. Autores

También se realiza el análisis de la señal en dBs, y de esta manera obtener sus valores de amplitud pico en dB, y de su frecuencia central.

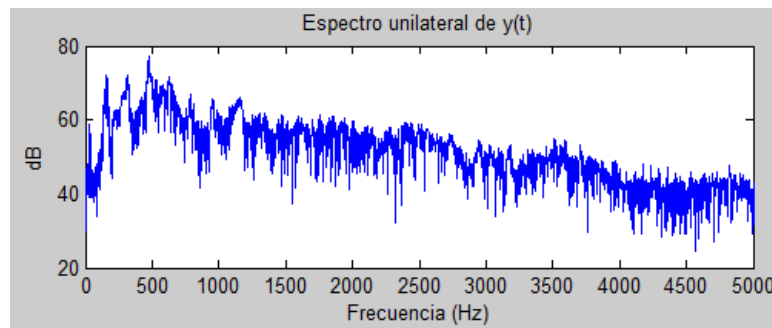


Figura 25. Espectro de frecuencia en dBs

Fuente. Autores

2.8.3.5 MUESTREO

Se realiza el muestreo de la señal para tomar sus valores discretos y realizar la gráfica del mismo.

```
n_bits1=input('ingrese el número de bits: ','s');
```

```
Fs1=input('ingrese el número de muestras: ','s');
```

```

n_bits=str2double(n_bits1);
Fs=str2double(Fs1);
disp('Procesando...');
[a,il]=size(y);
tiempo=a/fs;
Fs=round(Fs*tiempo);
figure(3)
plot(y);
wavplay(y,fs);
%% -----Muestreo-----
muestras=round(linspace(1,a,Fs));
senal_muestreada=zeros(1,Fs);
index=1;
for i=muestras
    senal_muestreada(index)=y(i);
    index=index+1;
end
stem(senal_muestreada);
disp('Muestreo realizado');

```

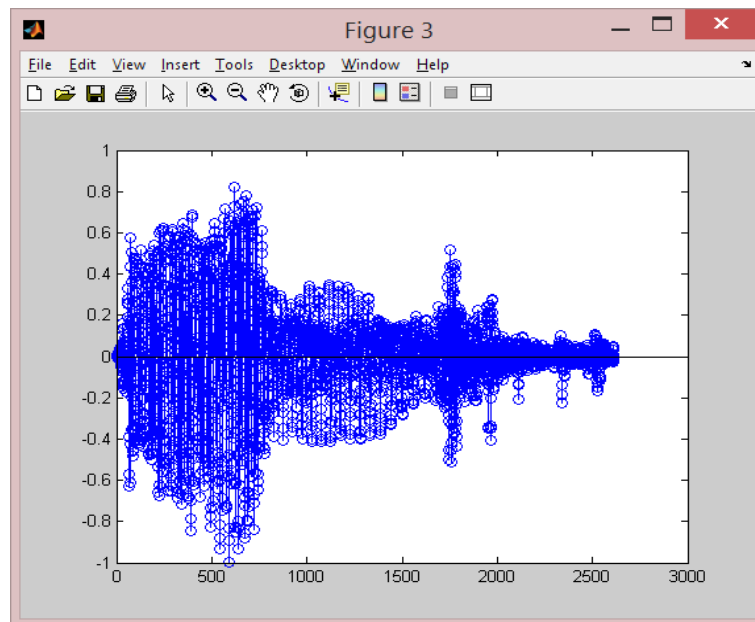


Figura 26. Señal muestreada

Fuente. Autores

2.8.3.6 CUANTIZACIÓN

Se realiza la cuantización la señal de entrada.

```
function
cuan=cuatizacion(y,n_bits,index,senal_muestreada,a,mues
tras,fs,frecuencia_acotada,amplitud)
%% -----Cuantización-----
vmax=max(y);
vmin=min(y);
lsb=(vmax-vmin)/(2^n_bits);
niveles=vmin:lsb:vmax;
particion=(vmin-(0.5*lsb)):lsb:(vmax+(0.5*lsb));
[il,indice_niv]=size(niveles);
% genera valores cuantizados de muestras
s_cuantizada=zeros(1,index-1);
%=====
=====
for i=1:index-1
    for j=1:indice_niv
% si el valor es <vmin entonces será igual a vmin
        if (senal_muestreada(i)<particion(1))
            s_cuantizada(i)=vmin;
            val_binairo=dec2bin(0,n_bits);

            rango_val=val_binario(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end
        if (senal_muestreada(i)>particion(end))
            s_cuantizada(i)=vmax;
            val_binairo=dec2bin(2^n_bits)-1,n_bits;
```

```

        rango_val=val_binario(:);
        rango_index=((i-1)*n_bits)+1;
        flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end

        if (senal_muestreada(i)>=particion(j))
            if (senal_muestreada(i)<particion(j+1))
                s_cuantizada(i)=niveles(j);
                %Al mismo tiempo la generación de
flujo binario para cada
                %valor cuantificado
                val_binairo=dec2bin(j-1,n_bits);

                rango_val=val_binairo(:);
                rango_index=((i-1)*n_bits)+1;
                flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
            end
        end
    end
end
[tam,il]=size(flujo);
figure(4)
stairs(s_cuantizada)
ylim([vmin vmax])
set(gca,'YTick',vmin:lsb:vmax)
disp('Cuantización realizada');

```

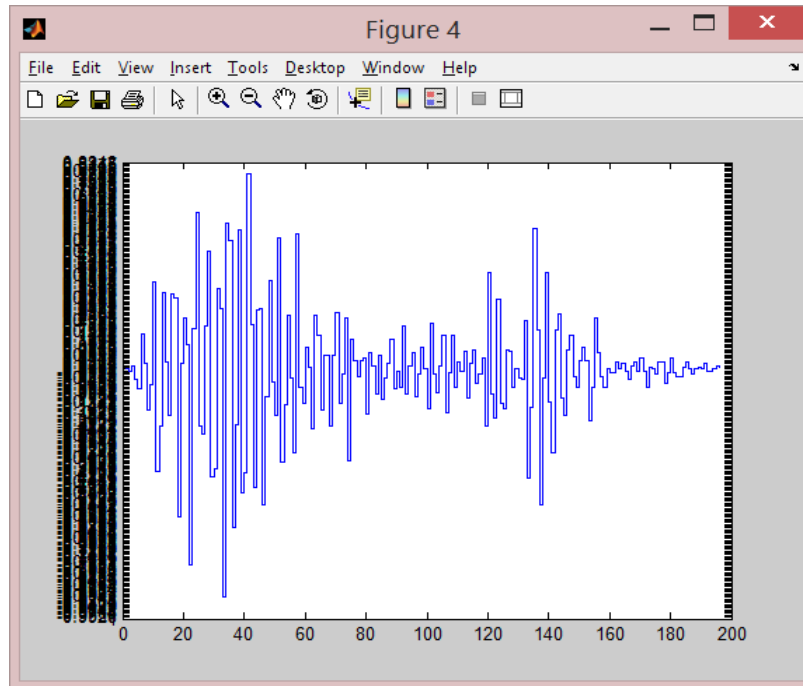


Figura 27. Señal cuantizada

Fuente. Autores

2.8.3.7 CODIFICACIÓN

Se obtiene el valor binario de la señal de acuerdo al número de bits y frecuencia de muestro seleccionados previamente.

```

val_cod=zeros(tam,1);
for i=1:tam
    val_cod(i,1)=str2double(flujo(i,1));
end
figure(2)
subplot(212)
stairs(val_cod');
axis([1 1000 -3 4]);
set(gca,'YTick',-3:1:4);
set(gca,'YTicklabel',{' ',' ','0','1',' ',' ',' '});

```

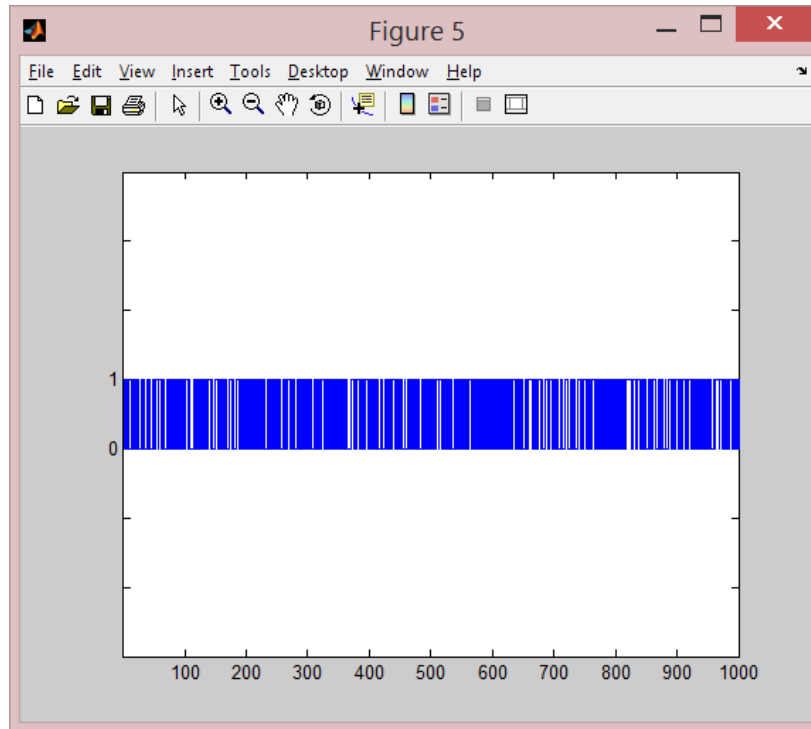


Figura 28. Señal codificada

Fuente. Autores

2.8.3.8 OBTENCION DE CARACTERISTICAS

Se realiza el cálculo de las características dentro de todo el proceso para la señal de entrada en tiempo real, y así realizar el proceso de comparación y clasificación.

```
function
[fs,mediana,Longitud,media,maximo,minimo,desviacion_est
andar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,
amplitud]=caracteristicas(y,fs)
% [y,fs]=wavread('USUARIO');
disp('Características de la voz');
pause;
sound(y,fs)
Fs=fs;
%% parámetros
```

```

mediana=median(y)%calcula la mediana
Longitud=length(y); %Longitud de y
media=mean(y)%calcula la media
maximo=max(y)%calcula el valor más alto
minimo=min(y)%calcula el valor más bajo
desviacion_estandar=std(y)%calcula la desviación
estándar
Varianza = var(y)
Curtosis = kurtosis(y) % cálculo de curtosis
Asimetria=skewness(y) % cálculo de la asimetría
%http://probestunalmzl.wikispaces.com/cap_1_p_11
% y= número total de muestras del vector
for i=1:(length(y)-1)
    var1=abs(y(i+1)-y(i));
end
V1=(1/((length(y)-1))*var1;
for i=1:(length(y)-1)
    var2=abs(y(i));
end
V2=(1/((length(y)-1))*var2;
Vp=V1/V2; %perturbación de la amplitud

for j=1:(length(y)-1)
    va1=abs(max(y(j+1))-max(y(j)));
end
V3=(1/((length(y)-1))*va1;
for j=1:(length(y)-1)
    va2=max(abs(y(j)));
end
V4=(1/((length(y)-1))*va2;
Sh=V3/V4 %cálculo de Shimer
%%
for k=1:(length(y)-1)

```



```

        v1=(y(k+1)-y(k))^2;
end
V5=sqrt((1/(length(y))))*v1);
for i=1:(length(y)-1)
    v2=abs(y(k));
end
V6=(1/(length(y)-1))*v2;
Cap=V5/V6 %perturbación de la amplitud
f=linspace(-Fs/2,Fs/2,length(y)); %Vector creado para
el eje de frecuencias
sa=fftshift(abs(fft(y))); %transformada rápida de
Fourier a la cual se le saca el valor absoluto
filal=sa.' % Organiza los datos en vector fila.
amp=max(sa); % Pico máximo
subplot(211)
plot(f,fftshift(abs(fft(y))))%reorganiza las salidas de
FFT, moviendo el componente de frecuencia cero hasta el
centro de la matriz.
%Es útil para la visualización de una transformada de
Fourier con el componente de frecuencia cero en el
centro del espectro.
w=find(sa==amp);%encuentra la posición del valor pico
mw=max(w); %halla el valor más alto
freq=f(mw)%halla la frecuencia central haciendo una
relación con la amplitud pico del espectro de
frecuencia
disp('Cálculo de frecuencia realizado');
%%
fs=Fs;
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiete potencia de 2 de la
long de y
Y = fft(y,NFFT)/L; % FFT de la señal

```

```

f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1))./4e-10);
amplitud=max(FDB);
w=find(FDB==amplitud);%encuentra la posición de
mw=max(w); %halla el valor más alto
frecuencia=f(mw)%halla la frecuencia central haciendo
una relación con la amplitud pico del espectro de
frecuencia

nam=input('Ingrese un nombre para almacenar las
características: ','s');
save
(nam,'fs','mediana','Longitud','media','maximo','minimo
','desviacion_estandar','Varianza','Curtosis','Asimetri
a','Vp','Sh','Cap','frecuencia','amplitud');
disp('Parámetros estadísticos obtenidos correctamente')
pause;
(nam,'fs','mediana','Longitud','media','maximo','minimo
','desviacion_estandar','Varianza','Curtosis','Asimetri
a','Vp','Sh','Cap','frecuencia','amplitud');
disp('Parámetros estadísticos obtenidos correctamente')
pause;

```

2.8.3.9 OBTENCION DE CARACTERISTICAS DE MEL

Se obtienen las características de los MFCCs para los coeficientes cepstrales de la grabación en tiempo real, y así poder realizar el proceso de clasificación de los estados emocionales realizando la comparación con los obtenidos en la base de datos para cada emoción.

```

function
[medianamfcc,Longitudmfcc,mediamfcc,maximomfcc,minimomf
cc,desviacion_estandarmfcc,Varianzarmfcc,Curtosismfcc,As

```

```

imetriamfcc,Vpmfcc,Shmfcc,Capmfcc]=caracteristicas_Mel(
MFCCs)
clc;
disp('Características de los parámetros MEL');
pause;
%% parametros
medianamfcc=median(MFCCs);%calcula la mediana
Longitudmfcc=length(MFCCs); %Longitud de MFCCs
mediamfcc=mean(MFCCs);%calcula la media
maximomfcc=max(MFCCs);%calcula el valor más alto
minimomfcc=min(MFCCs);%calcula el valor más bajo
desviacion_estandarmfcc=std(MFCCs);%calcula la
desviación estándar
Varianzmfcc = var(MFCCs);
Curtosismfcc = kurtosis(MFCCs); % Cálculo de curtosis
Asimetriamfcc=skewness(MFCCs); % Cálculo de la
asimetría
%http://probestunalmzl.wikispaces.com/cap_1_p_11
% MFCCs= numero total de muestras del vector
for i=1:(length(MFCCs)-1)
    var1=abs(MFCCs(i+1)-MFCCs(i));
end
V1=(1/((length(MFCCs))-1))*var1;
for i=1:(length(MFCCs)-1)
    var2=abs(MFCCs(i));
end
V2=(1/((length(MFCCs))-1))*var2;
Vpmfcc=V1/V2; %perturbación de la amplitud

for j=1:(length(MFCCs)-1)
    val=abs(max(MFCCs(j+1))-max(MFCCs(j)));
end
V3=(1/((length(MFCCs))-1))*val;

```

```

for j=1:(length(MFCCs)-1)
    va2=max(abs(MFCCs(j)));
end
V4=(1/(length(MFCCs)-1))*va2;
Shmfcc=V3/V4; %cálculo de Shimer

for k=1:(length(MFCCs)-1)
    v1=(MFCCs(k+1)-MFCCs(k))^2;
end
V5=sqrt((1/(length(MFCCs)))*v1);
for i=1:(length(MFCCs)-1)
    v2=abs(MFCCs(k));
end
V6=(1/(length(MFCCs)-1))*v2;
Capmfcc=V5/V6 %perturbación de la amplitud
%%
nam=input('Ingrese un nombre para almacenar las
características: ','s');
save
(nam,'medianamfcc','Longitudmfcc','mediamfcc','maximomf
cc','minimomfcc','desviacion_estandarmfcc','Varianzamfc
c','Curtosismfcc','Asimetriamfcc','Vpmfcc','Shmfcc','Ca
pmfcc');
disp('Parámetros estadísticos obtenidos correctamente')
pause;

```

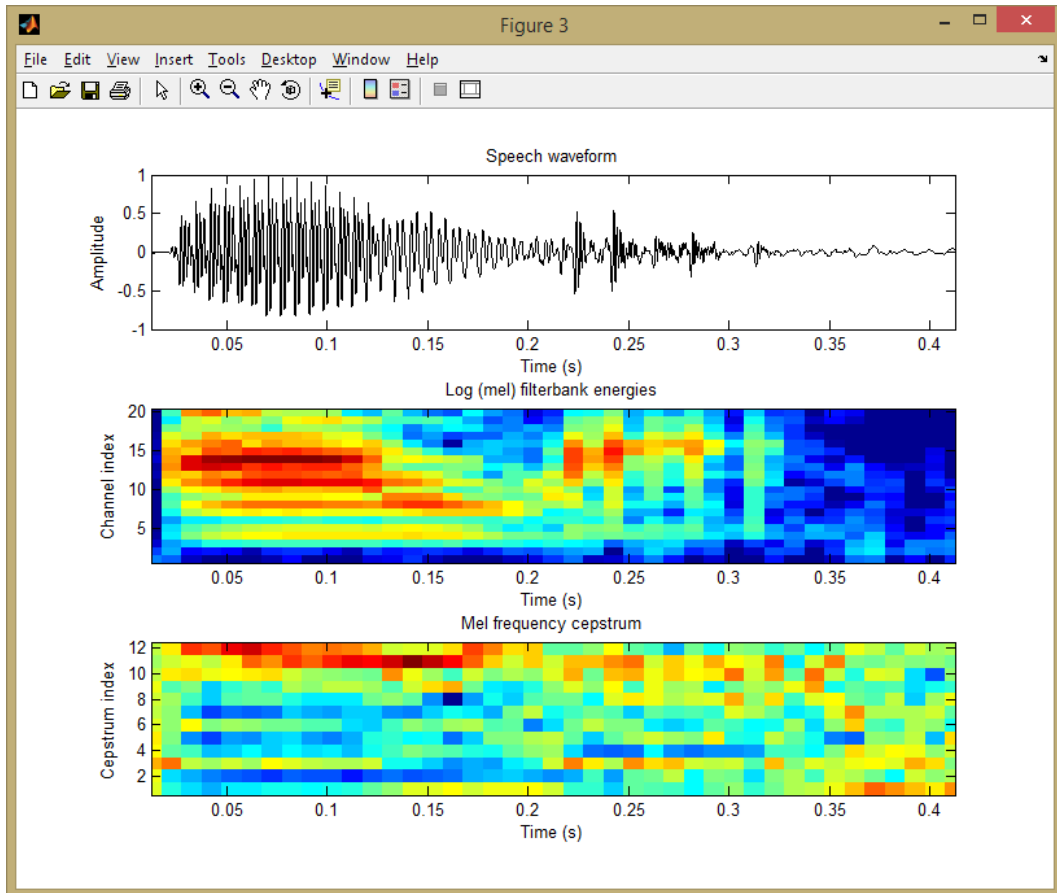


Figura 29. A) Señal en función del tiempo. B) Índice del canal en el tiempo. C) Índices cepstrales.

Fuente: Autores

2.8.3.10 PROCESO

En esta etapa se realiza el procesamiento, reconocimiento y clasificación de los estados emocionales. Con una señal de entrada se realizan comparaciones con la base de datos con el fin de encontrar la que tenga características similares con la señal de entrada.

```
clc;
load
('Target_1.mat','fs','mediana','Longitud','media','maximo','minimo','desviacion_estandar','Varianza','Curtosis','Asimetria','Vp','Sh','Cap','frecuencia','amplitud');
```

```

a=[fs,mediana,Longitud,media,maximo,minimo,desviacion_
standar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuenci
a,amplitud]
pause
load
('Target_2.mat','fs','mediana','Longitud','media','maxi
mo','minimo','desviacion_estandar','Varianza','Curtosis
','Asimetria','Vp','Sh','Cap','frecuencia','amplitud')
b=[fs,mediana,Longitud,media,maximo,minimo,desviacion_
standar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuenci
a,amplitud];
pause
load
('Target_3.mat','fs','mediana','Longitud','media','maxi
mo','minimo','desviacion_estandar','Varianza','Curtosis
','Asimetria','Vp','Sh','Cap','frecuencia','amplitud')
c=[fs,mediana,Longitud,media,maximo,minimo,desviacion_
standar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuenci
a,amplitud];
pause
load
('Target_4.mat','fs','mediana','Longitud','media','maxi
mo','minimo','desviacion_estandar','Varianza','Curtosis
','Asimetria','Vp','Sh','Cap','frecuencia','amplitud')
d=[fs,mediana,Longitud,media,maximo,minimo,desviacion_
standar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuenci
a,amplitud];
pause
load
('INPUT.mat','fs','mediana','Longitud','media','maximo'
,'minimo','desviacion_estandar','Varianza','Curtosis','
Asimetria','Vp','Sh','Cap','frecuencia','amplitud');

```

```

s=[fs,mediana,Longitud,media,maximo,minimo,desviacion_e
standar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuenci
a,amplitud];
pause
error(1)=mean(abs(a-s));
error(2)=mean(abs(b-s));
error(3)=mean(abs(c-s));
error(4)=mean(abs(d-s));

min_error=min(error);

display(error);
disp(min_error);

if (min_error == error(1))
    display('Feliz')
end
if (min_error == error(2))
    display('Tristeza')
end
if (min_error == error(3))
    display('Miedo')
end
if (min_error == error(4))
    display('Sorpresa')
end

```

2.9 COMPROBACIÓN DE LA HIPÓTESIS

La comprobación de la hipótesis es un procedimiento basado en la evidencia de muestras y la teoría de probabilidad; se emplea para determinar si la hipótesis es una afirmación razonable y para inferir que cierta condición es válida para toda la población. En concreto, el procedimiento que da forma a cualquier prueba de

hipótesis, se expone, que se acomete siguiendo estos pasos fundamentales: planteamiento de la hipótesis nula y de la alternativa.

La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que "no hay efecto" o "no hay diferencia".

La hipótesis alternativa es el enunciado que se desea poder concluir que es verdadero.

Con base en los datos de la muestra, la prueba determina si se debe rechazar la hipótesis nula. Para tomar la decisión se utiliza un valor p. Si el valor p es menor que o igual al nivel de significancia, que es un punto de corte que usted define, entonces puede rechazar la hipótesis nula.

2.10 PLANTEAMIENTO DE LA HIPÓTESIS ESTADÍSTICA

Hipótesis nula (H₀): El diseño de un software no permitirá la clasificación de estados emocionales por medio del reconocimiento de tonos de voz y gramáticas.

Hipótesis específica (H₁): El diseño de un software permitirá la clasificación de estados emocionales por medio del reconocimiento de tonos de voz y gramáticas.

H₀: $n_1 \geq n_2$

H₁: $n_1 < n_2$

2.10.1 ESTABLECIMIENTO DEL NIVEL DE SIGNIFICANCIA.

Se realizaron las pruebas con un 95% de confiabilidad, es decir, se trabajó con un nivel de significancia de $\alpha=0.05$.

2.10.2 DETERMINACIÓN DEL VALOR ESTADÍSTICO DE PRUEBA.

Rechaza la **H₀** si la $Z_c > 1.64$

Donde 1.64 es el valor de Z en un ensayo a una cola con un nivel de significación 0.05, y Z_c es el valor calculado de Z que se obtiene aplicando la fórmula:

$$Z_c = \frac{P_1 - P_2}{\sqrt{\frac{p_1q_1}{n} + \frac{p_1q_1}{n_1}}}$$

Para lo cual se requirió tomar un número de muestras de las pruebas con el sistema.

2.11 PRUEBAS DE FUNCIONAMIENTO DEL SISTEMA

2.11.1 ESEENARIO 1

Categoría	Grupo A		Grupo B	
	Ni	%	ni	%
Miedo	3	15	10	50
Alegría	5	25	7	35
Tristeza	5	25	2	10
Sorpresa	7	35	1	5
TOTAL	20	100	20	100

Tabla 3. Diferencias de los escenarios de prueba

Fuente. Autores

Índice de significancia de los estados emocionales del escenario 1

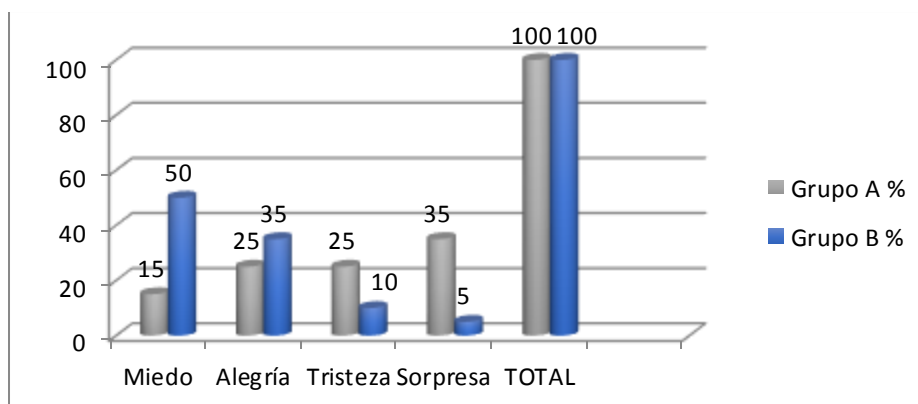


Figura 30. Porcentaje de Pruebas del escenario1

Fuente. Autores

2.11.2 ESCENARIO 2

Categoría	Grupo A		Grupo B	
	Ni	%	ni	%
Miedo	3	15	9	45
Alegría	5	25	8	40
Tristeza	6	30	1	5
Sorpresa	6	30	2	10
TOTAL	20	100	20	100

Tabla 4. Diferencias de los escenarios de prueba

Fuente. Autores

Índice de significancia de los estados emocionales del escenario 2

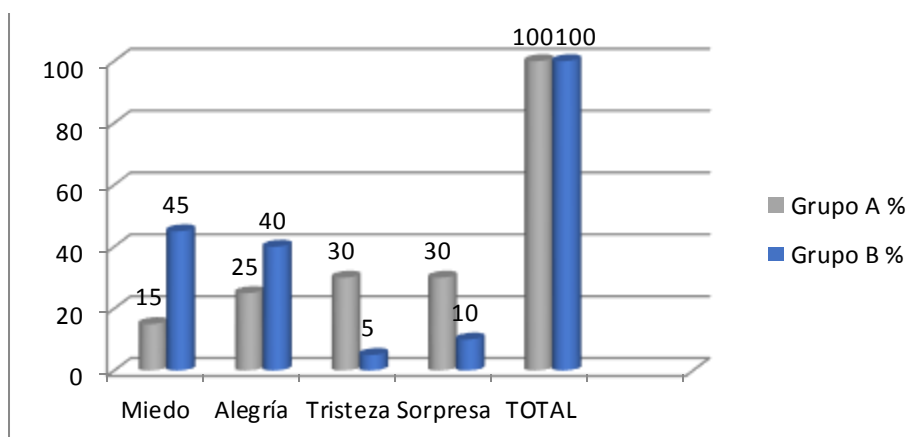


Figura 31. Porcentaje de Pruebas del escenario 2

Fuente. Autores

2.11.3 ESCENARIO 3

Categoría	Grupo A		Grupo B	
	Ni	%	ni	%
Miedo	2	10	9	45
Alegría	2	10	7	35
Tristeza	8	40	2	10
Sorpresa	8	40	2	10
TOTAL	20	100	20	100

Tabla 5. Diferencias de los escenarios de prueba

Fuente. Autores

Índice de significancia de los estados emocionales del escenario 3

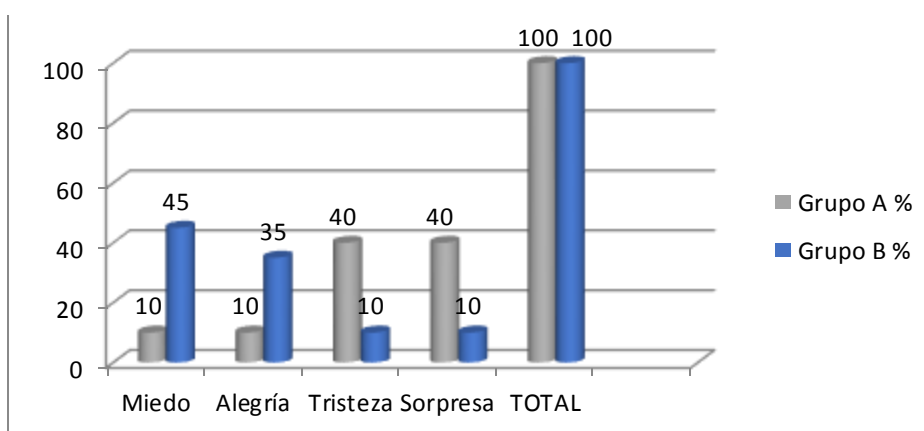


Figura 32. Porcentaje de Pruebas del escenario 3

Fuente. Autores

2.11.4 ESCENARIO 4

Categoría	Grupo A		Grupo B	
	Ni	%	ni	%
Miedo	3	15	9	45
Alegría	4	20	9	45
Tristeza	6	30	1	5
Sorpresa	7	35	1	5
TOTAL	20	100	20	100

Tabla 6. Diferencias de los escenarios de prueba

Fuente. Autores

Índice de significancia de los estados emocionales del escenario 4

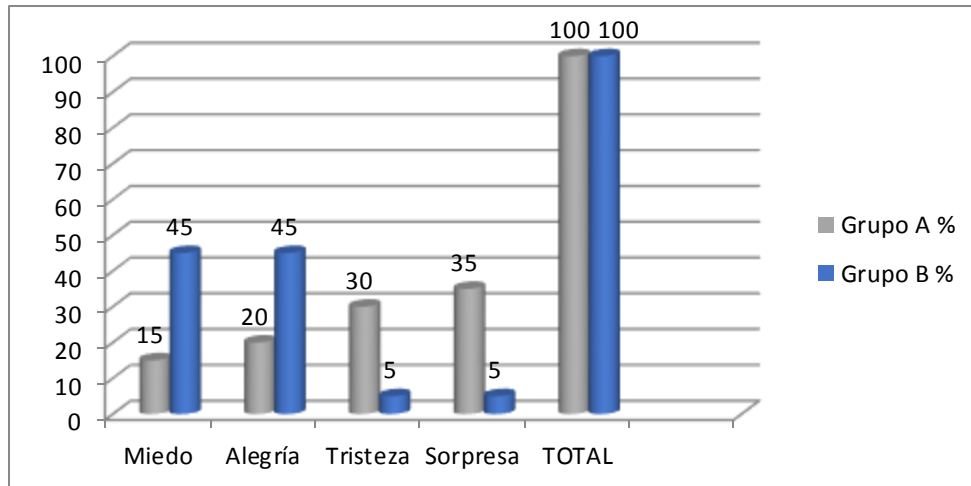


Figura 33. Porcentaje de Pruebas del escenario 4

Fuente. Autores

2.11.5 ESCENARIO 5

Categoría	Grupo A		Grupo B	
	Ni	%	ni	%
Miedo	1	5	9	45
Alegría	4	20	8	40
Tristeza	7	35	1	5
Sorpresa	8	40	2	10
TOTAL	20	100	20	100

Tabla 7. Diferencias de los escenarios de prueba

Fuente. Autores

Índice de significancia de los estados emocionales del escenario 5

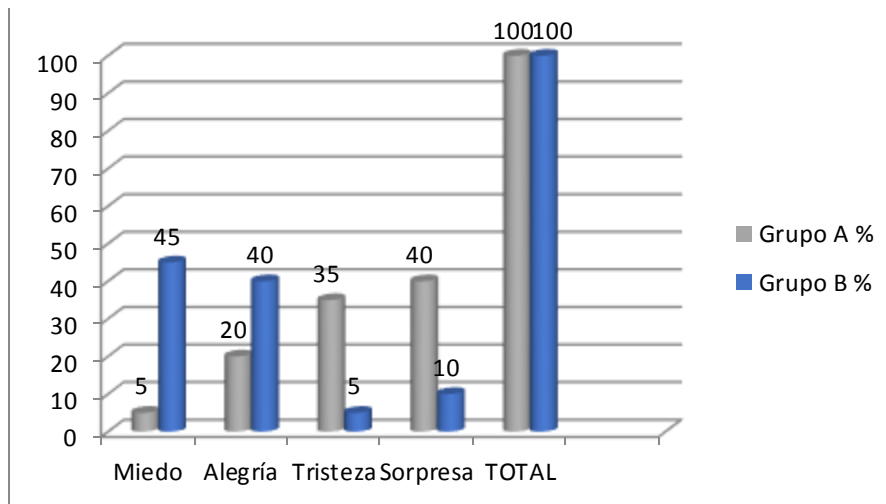


Figura 34. Porcentaje de Pruebas del escenario 5

Fuente. Autores

2.11.6 ANÁLISIS DEL RESULTADO DEL GRUPO A

N°	Decisión Pregunta	Miedo	Alegría	Tristeza	Sorpresa	TOTAL
1	Escenario 1	3	5	5	7	20
2	Escenario 2	3	5	6	6	20
3	Escenario 3	2	2	8	8	20
4	Escenario 4	3	4	6	7	20
5	Escenario 5	1	4	7	8	20
TOTAL		12	20	32	36	100

Tabla 8. Datos obtenidos en la aplicación del software a los estudiantes

Fuente. Autores

2.11.7 ANÁLISIS DEL RESULTADO DEL GRUPO B

N°	Decisión Pregunta	Miedo	Alegría	Tristeza	Sorpresa	TOTAL
1	Escenario 1	10	7	2	1	20
2	Escenario 2	9	8	1	2	20
3	Escenario 3	9	7	2	2	20
4	Escenario 4	9	9	1	1	20
5	Escenario 5	9	8	1	2	20
TOTAL		46	39	7	8	100

Tabla 9. Datos obtenidos en la aplicación del software a los estudiantes del grupo B

Fuente. Autores

2.11.8 PORCENTAJE DE RESULTADOS DE LOS GRUPOS DE PRUEBA

Grupo Categoría	GRUPO A		GRUPO B	
	Ni	%	Ni	%
Miedo	12	12	46	46
Alegría	20	20	39	39
Tristeza	32	32	7	7
Sorpresa	36	36	8	8
TOTAL	100	100	100	100

Tabla 10. Porcentajes de prueba

Fuente. Autores

2.11.9 VALOR ESTADÍSTICO DE LOS ESTADOS EMOCIONALES DEL GRUPO A

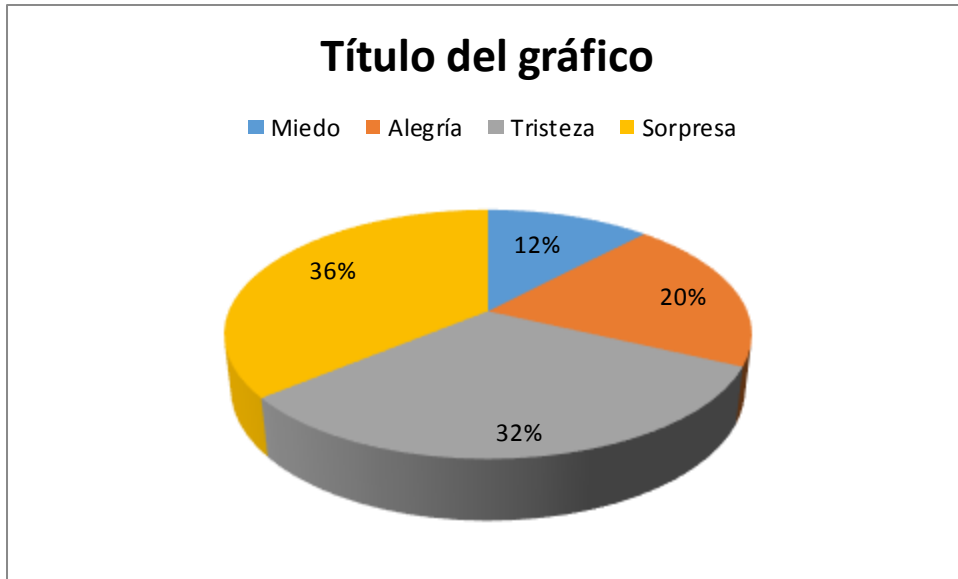


Figura 35. Estados emocionales del grupo A

Fuente. Autores

2.11.10 VALOR ESTADÍSTICO DE LOS ESTADOS EMOCIONALES DEL GRUPO B

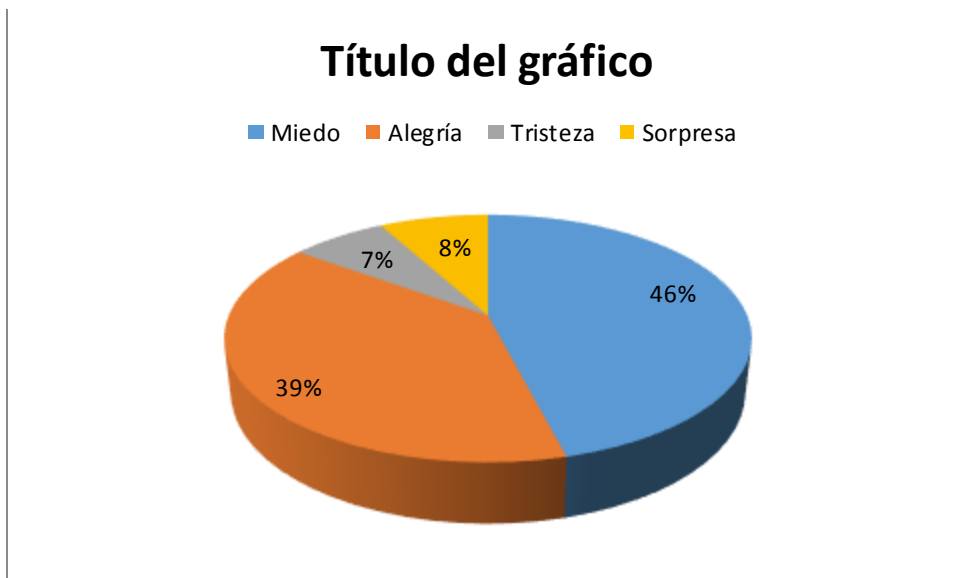


Figura 36. Porcentajes de prueba

Fuente. Autores

El análisis que se obtuvo en la aplicación del software establece un porcentaje de estados emocionales tanto para el grupo A y B y se tomó en consideración para el análisis simplemente dos estados emocionales para cada grupo de prueba en los estudiantes.

Aplicación de la fórmula:

$$Z_c = \frac{P_1 - P_2}{\sqrt{\frac{p_1q_1}{n} + \frac{p_2q_2}{n_2}}}$$

La simbología que se utilizó es la siguiente:

P1: proporción muestral 1

P2: proporción muestral 2

n1, n2: tamaño de la muestra

q1= pc-1

q2= ps-1

2.11.11 CALCULO ESTADÍSTICO

N° ESCENARIOS	GRUPO A				GRUPO B			
	Miedo	Alegría	Tristeza	Sorpresa	Miedo	Alegría	Tristeza	Sorpresa
	%	%	%	%	%	%	%	%
1	15	25	25	35	50	35	10	5
2	15	25	30	30	45	40	5	10
3	10	10	40	40	45	35	10	10
4	15	20	30	35	45	45	5	5
5	5	20	35	40	45	40	5	10
Total	50	100	160	180	230	195	35	40
PROMEDIO	10	20	32	36	46	39	7	8

Tabla 11. Variaciones de Estados Totales de los grupos de prueba

Fuente. Autores

2.11.12 VARIACIONES TOTALES DE LOS GRUPOS DE PRUEBA

Miedo y Alegría	Grupo A	30
	Grupo B	85

Tabla 12. Valor Final de prueba

Fuente. Autores

$$P1 = 0,3$$

$$P2 = 0,85$$

$$q1 = 1 - 0,3 = 0,7$$

$$q2 = 1 - 0,85 = 0,15$$

$$n1 = 10$$

$$n2 = 10$$

$$Z_c = \frac{P2 - P1}{\sqrt{\frac{p2q2}{n2} + \frac{p1q1}{n1}}}$$

$$z_c = \frac{0,85 - 0,3}{\sqrt{\frac{(0,85 * 0,15)}{10} + \frac{(0,3 * 0,7)}{10}}}$$

$$z_c = \frac{0,55}{\sqrt{0.01275 + 0.021}}$$

$$z_c = \frac{0,55}{\sqrt{0.03375}}$$

$$z_c = \frac{0,55}{0.1837} = 2,99$$

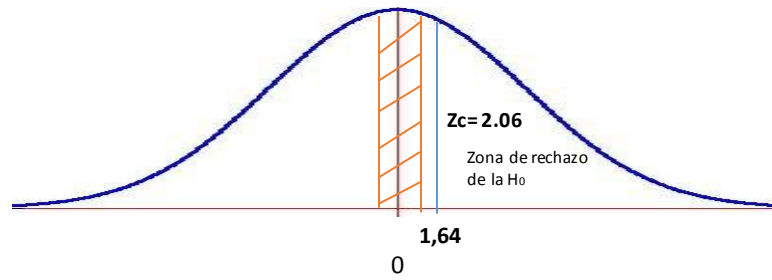


Figura 37. Resultado de Comparación

Fuente. Autores

Interpretación: Como el valor de Z_c es igual a 2,99 se encuentra fuera del área de aceptación de la hipótesis nula por lo tanto se acepta la hipótesis específica, esto significa que el software para la “Clasificación de los estados emocionales mediante tono de voz y gratica a través de la técnica de observación desarrolla la capacidad de detectar los estados emocionales de los estudiantes de la Universidad Nacional De Chimborazo.

CAPÍTULO III

3 ANALISIS DE RESULTADOS

El estudio sobre la clasificación de los estados emocionales permitió realizar un análisis de tipo comparativo entre las diferentes características de las señales de voz, en función de tiempo y frecuencia, utilizando el software MATLAB Versión 7.5.0.342 (R2007b) para la elaboración del proyecto.

Los algoritmos para el análisis y procesamiento de las señales de voz fueron desarrollados para minimizar el margen de error y, obtener una mayor precisión en el reconocimiento de los estados emocionales, en cualquier entorno, siendo este el principal inconveniente, ya que es necesaria una óptima grabación de la voz para mejores resultados.

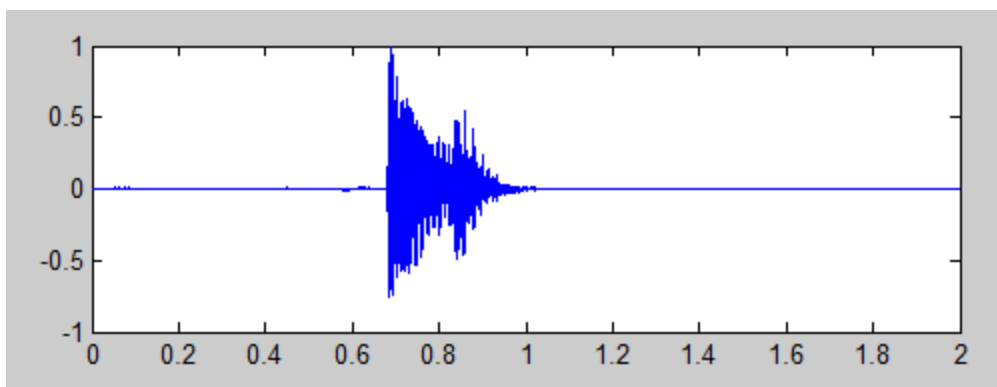


Figura 38. Señal de voz de entrada.

Fuente. Autores

Para el procesamiento de la señal de audio obtenida es necesario utilizar únicamente la porción de información útil y recortar la señal utilizando un algoritmo para eliminar los momentos de silencio segmentando la señal y basándose en la teoría las ventanas de Haming para segmentación.

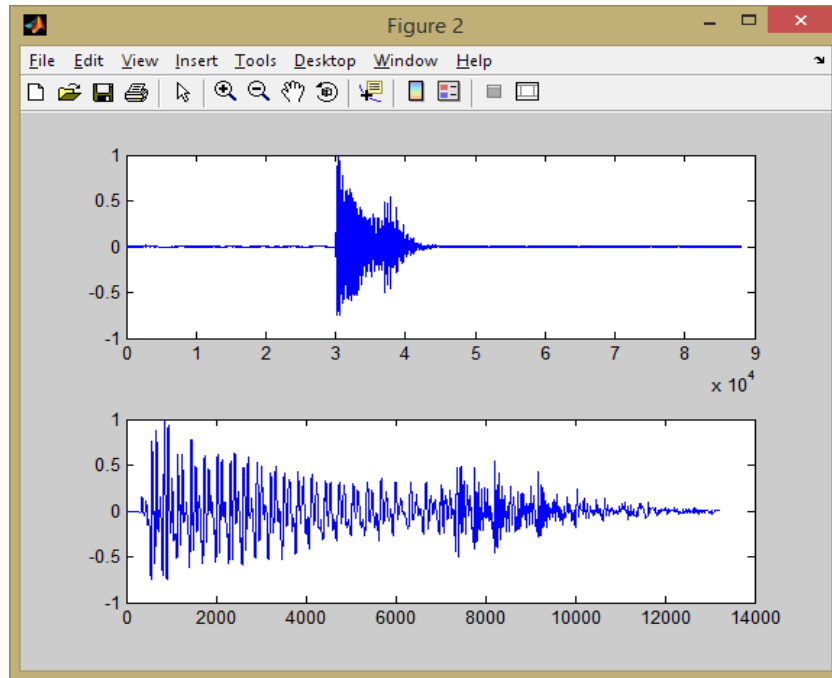


Figura 39. Eliminación de silencios

Fuente. Autores

Esta señal de información útil es procesada y se realiza la conversión analógica a digital, para así obtener sus características necesarias.

MUESTREO

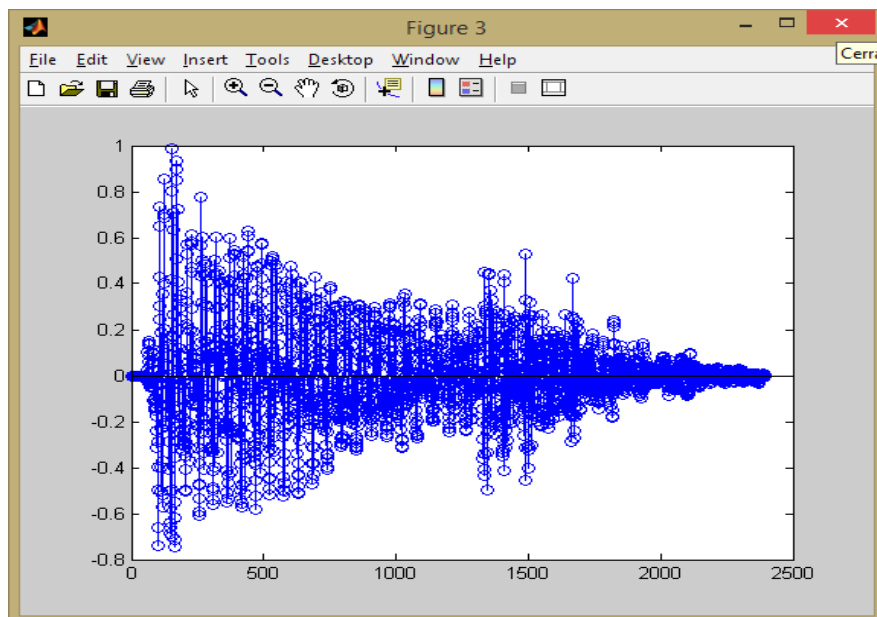


Figura 40. Señal de voz muestreada.

Fuente. Autores

CUANTIZACION

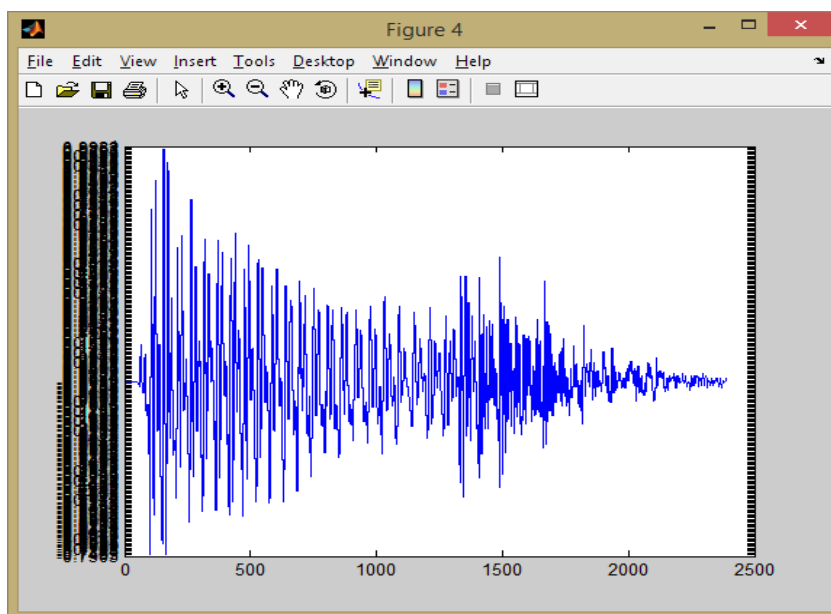


Figura 41. Señal de voz cuantizada.

Fuente. Autores

CODIFICACION

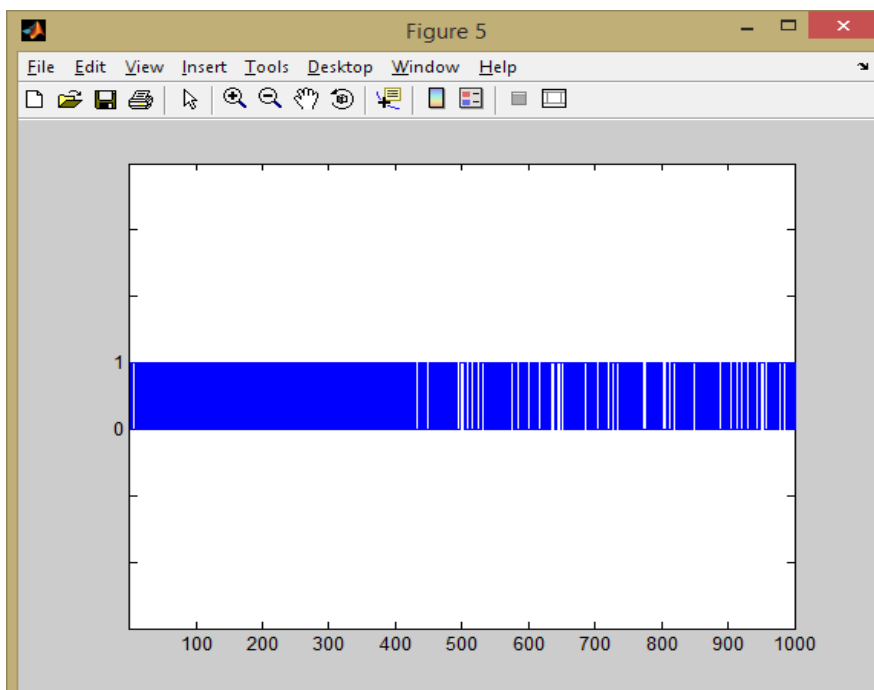


Figura 42. Señal codificada.

Fuente. Autores

También es necesario obtener las características de la señal en el dominio de la frecuencia, para así hallar valores de la frecuencia y amplitud de la señal

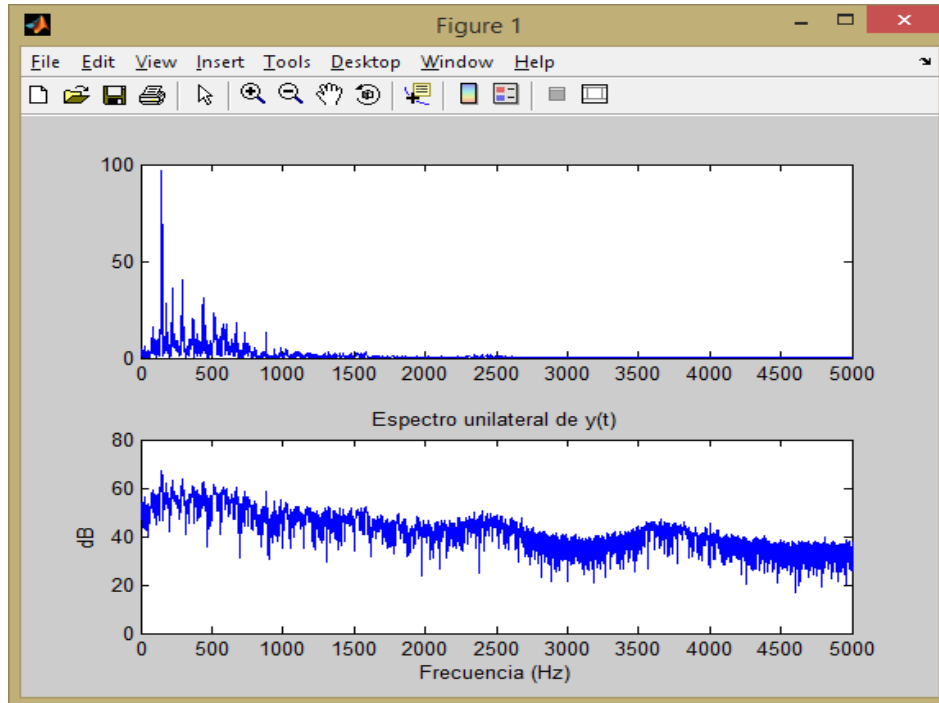


Figura 43. Espectro de frecuencia.

Fuente. Autores

Una vez extraídas las características de la señal, esta es comparada con las obtenidas previamente, clasificadas y almacenadas en una base de datos de MATLAB, para luego realizar la comparación y clasificación de los estados emocionales buscando las que coincidan con el menor número de errores en el análisis y así el programa procede a la clasificación de los estados emocionales de las personas.

CAPÍTULO IV

4 DISCUSIÓN

- ✓ El diseño del software mediante el programa matlab permite la creación de un reconocedor de estados emocionales de las personas, mediante el análisis de voz, que clasifica según la voz, en qué estado se encuentra los usuarios al momento que lo utiliza.
- ✓ El reconocimiento de los estados emocionales de las personas, se realizó mediante un amplio proceso de análisis de la voz de las personas, con el uso del software, que facilito la obtención de resultados eficientes.
- ✓ El software permite la iteración humano-computador, para el análisis de reconocimiento de estados emocionales se crea una base de datos la cual va ser procesada y analizada por el sistema.
- ✓ El ruido son señales que perturban la transmisión y procesamiento de señales, el ruido se vuelve un problema creciente que afecta al desarrollo del software. Para que las señales no tengan ningún tipo de ruido se empleó una un filtro Butterworth pasa banda con la finalidad de obtener una señal dentro del rango de frecuencias audibles para el oído humano.
- ✓ Culminando el desarrollo del software de reconocimiento de estados emocionales de las personas mediante la voz, se logró cumplir todas las especificaciones expuestas.

CAPÍTULO V

5 CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- ✓ La utilización del programa Matlab fue esencial para el desarrollo del software para el análisis y clasificación de los estados emocionales de las personas.
- ✓ Los resultados obtenidos dependen de las condiciones de grabación, y del ambiente, para obtener un menor margen de error en el reconocimiento de cada estado emocional.
- ✓ El emplear la base de datos genera resultados efectivos de reconocimiento de estados emocionales, gracias a su diseño el sistema puede ser modificado o ampliado para incorporar nuevas características de clasificación.
- ✓ El software de reconocimiento de estados emocionales generado en Matlab analiza las variaciones en las señales de voz, además también realiza el análisis de los MFCCs utilizando los coeficientes cepstrales de frecuencia de Mel, las cuales depende de cada estado emocional, se busca identificar estas variaciones para así lograr identificar los estados emocionales.
- ✓ Una vez finalizado el proyecto, se realizó pruebas y, mediante el análisis de los estados emocionales de las personas, se verificó el correcto funcionamiento de las clasificaciones generadas por el software.

5.2 RECOMENDACIONES

Luego de haber realizado las conclusiones de la investigación se propone las siguientes recomendaciones:

- ✓ Es necesario ampliar la información acerca del reconocimiento de los estados emocionales de las personas, debido a que la información existente es muy limitada y está más enfocadas a lo que respecta a reconocimiento de VOZ.
- ✓ La información respecto a las diferencias en tono y frecuencia de las voces masculina y femenina son muy escasas, lo cual causo dificultades al momento de la elaboración del software.
- ✓ Al momento de ejecutar el programa se debe tener en consideración que el margen de error del programa depende del ambiente en el que se está trabajando.
- ✓ Para mayor beneficio de las personas el software debe detectar correctamente el estado emocional en el que se encuentra al momento que se hace la iteración humano – computador en tiempo real.
- ✓ Es recomendable que el software pase por un proceso de entrenamiento para verificar que se encuentre en perfecto funcionamiento.

CAPÍTULO VI

6 PROPUESTA

6.1 TITULO DE LA PROPUESTA

“CLASIFICACIÓN DE ESTADOS EMOCIONALES POR MEDIO DEL RECONOCIMIENTO DE TONOS DE VOZ Y GRAMÁTICAS”

6.2 INTRODUCCIÓN

El reconocimiento de las emociones es un proceso mediante el cual la señal de voz será procesada y analizada a través de un software. Esta tecnología se sitúa dentro del marco más general del procesamiento de la voz y otras tecnologías como síntesis de voz, codificación de voz. En general, estas tecnologías permiten el análisis mediante una máquina que posee habilidad humana de escuchar e identificar los estados emocionales de la personas.

En la actualidad los sistemas de reconocimiento de estados emocionales tienen un gran desarrollo e importancia en la vida cotidiana.

Los diferentes métodos que existen para el reconocimiento de estados de ánimo, analizan las señales de la voz en función de la frecuencia, y debe atravesar diversas etapas para procesar y analizar, utilizando un software diseñado para este propósito.

6.3 OBJETIVOS

6.3.1 OBJETIVO GENERAL

- ✓ Construir un software para detectar el estado de ánimo de las personas utilizando la voz.

6.3.2 OBJETIVOS ESPECIFICOS

- ✓ Identificar los diferentes estados de ánimos de las personas utilizando la voz.
- ✓ Diseñar el algoritmo de clasificación de estados emocionales

- ✓ Diseñar un software para identificar los estados de ánimos de la persona utilizando la voz.

6.4 FUNDAMENTACIÓN CIENTÍFICA

El diseño de un software para el reconocimiento de estados emocionales de las personas mediante los tonos de voz está formado por procesos de entrada de voz, pre-procesamiento, reconocimiento, comunicación y la aplicación del software.

El procesamiento de voz funciona con un transductor, el cual convierte la voz de señal analógica a digitales, el software realiza procesos de análisis y comprobaciones de los estados, los datos adquiridos son comparados con los obtenidos previamente y almacenados en una base de datos, generando comparaciones entre cada estado y así decidir el estado emocional que tiene la persona.

6.5 DESCRIPCIÓN DE LA PROPUESTA

El presente proyecto se pone a beneficio de todas las personas que requieran saber en qué estado emocional se encuentran, tiene como finalidad mejorar su estilo de vida.

El primer paso consiste en la creación de una base de datos con la voz de las personas, en la que se almacenara las características de las voces de acuerdo a cada emoción, facilitando así que el software detecte de manera correcta en qué estado se encuentra cada persona.

Luego esta parte del reconocimiento, es decir la identificación de cada estado emocional, el resultado obtenido de las actividades realizadas de procesamiento de voz y los métodos de procesamiento digital el algoritmo utilizado en el programa permitió almacenar información necesaria para su funcionamiento.

Finalmente se realiza la detección de los estados emocionales de las personas mediante el desarrollo del software, logrando que el usuario identifique gracias al software los diferentes estados emocionales.

6.6 DISEÑO ORGANIZACIONAL.

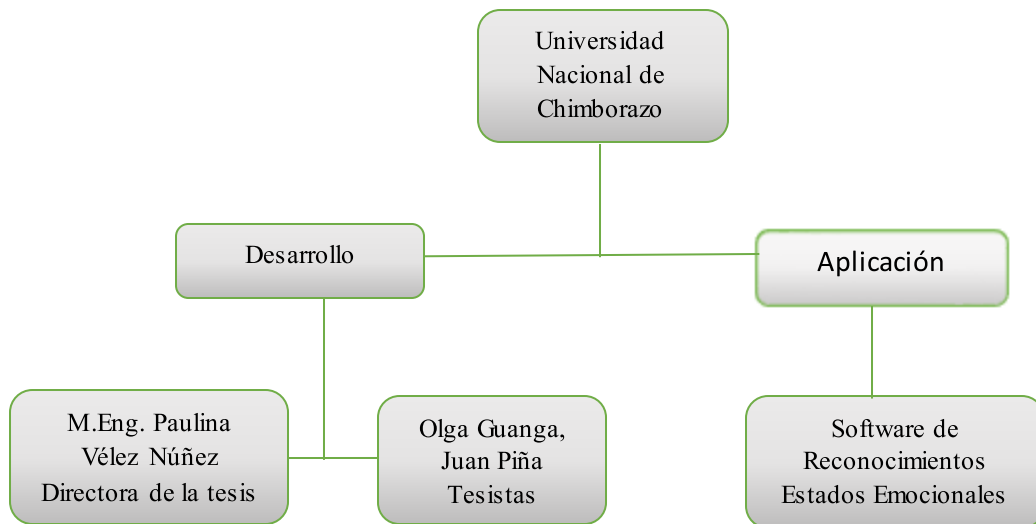


Figura 44. Esquema Organizacional

Fuente. Autores

6.7 MONITOREO Y EVALUACIÓN DE LA PROPUESTA

El monitoreo y evaluación de la propuesta se la realizara a través de pruebas a un gran número de personas, para determinar si el la eficiencia del software lograr su relevancia de reconocimientos.

Para una correcta evaluación del software se documentara cada prueba realizada, de manera que, al revisar el análisis y la confiabilidad de funcionamiento del que existe, se pueda establecer las algunas limitaciones que presenta y poder corregirlas.

CAPÍTULO VII

7 BIBLIOGRAFÍA

- A., A. A. (2007). *Manual de Oratoria*. Electrónica Gratuita.
- Aguilar Torres, G., & Martínez Mascorro, M. (2012). *Mel-frequency cepstral coefficients, MFCC*. México.: INGENIUS.
- Alzina, R. B. (2003). *TEORIA DE LAS EMOCIONES*. Madrid: Revista de Investigación Educativa.
- Antón, E. R. (2001). *El tono de la voz masculina y femenina en los informativos radiofónicos: un análisis comparativo*. España.
- Arias, F. G. (2006). *Introducción a la Metodología Científica*. Caracas: Episteme.
- Daniel, V. (1988). *Fonética i Fonología*. Madrid: Bibliografía Internacional.
- Ekma, P. (1992). *Un argumento de la Emociones Básicas*. California: San Francisco, U.S.A.
- Ekman, P. (1992). *An Argument For Basic Emotions*. California: San Francisco .
- Farfán, C. A. (2013). LA VOZ HUMANA Y SU CLASIFICACION. *TECO ESUFA*, 4.
- Faúndez Zanuy, M. (2000). *Tratamiento Digital de Voz e Imagen con Aplicación a la*. Marcombo.
- Faúndez, M. (2000). *TRATAMIENTO DIGITAL*. México: Marcombo.
- Harriton, J. y. (2005). *Techniques in Speech Acoustics*. KluwerAcademicPublishers.
- Kai-Fu, L. (1990). *AutomaticSpeechRecognition*. Taiwan.
- M, C. U. (2008). *La Fonoaudiología de Voz y Audición*. México.
- Miyara, F. (2005). *La Voz Humana*. Bogotá: UPOLI, Universidad Politécnica de Nicaragua.
- Pelton, G. E. (1992). *Voice Processng*. Singapore: MacGraw.
- Pereira, A. (2003). *Procesado Digital de la Señal Sonora Utilizando MATLAB*. LULU.
- REEVE, J. (2002). *Motivación y Emoción* . México: McGRAW.HILL/INTERAMERICA EDITORES. .
- Salias, J. (2005). *INFLUENCIA DE LA EMOCIÓN EN LA VOZ*. Atlántico.
- Schroeder, M. R. (1999). *Computer Speech Compression*. New York: Springer.

- Vivas, M. (2007). *Educar las Emociones*. Venezuela: Producción Editorial C.A.
- Vives, D. (1993). *FONETICA Y FONOLIGIA*. MADRID.

CAPÍTULO VIII

8 ANEXOS

```
clc;clear all;
disp ('=====');
disp ('=          Escoja una opción          =');
disp ('=====');
opcion=input('1-Crear Base de Datos \n2-Clasificar
Estados (Hombre) \n3-Clasificar Estados (Mujer) \n4-
Salir\n');
clc;
switch opcion
%%
    case 1,
        [y,fs]=grabadora;
        nom=guardar(y,fs);
        [y,fs]=cargar(y,nom);
        L=longitud(y);
        [x3,ti]=normalizar(y,fs);
        [senal_filtrada,filtb,filta]=pasabajos_butterworth(x3,fs);
        k=silencio(senal_filtrada,fs,nom);
        fourier(k,fs,nom);
        [frecuencia_acotada,amplitud]=frecuencia_C(k,fs);
        [fs,mediana,Longitud,media,maximo,minimo,desviacion_estandar,
        Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,amplitud]=
        caracteristicas(k,fs);
        muestreo(k,fs,frecuencia_acotada,amplitud);
        Base_de_datos;
        clc;
        %%
    case 2 ,
        [y,fs]=grabadora;
        nom=guardar(y,fs);
```

```

[y,fs]=cargar(y,nom);
L=longitud(y);
[x3,ti]=normalizar(y,fs);
[senal_filtrada,filtb,filta]=pasabajos_butterworth(x3,fs);
k=silencio(senal_filtrada,fs,nom);
fourier(k,fs,nom);
[frecuencia_acotada,amplitud]=frecuencia_C(k,fs);
[ MFCCs, FBEs, frames ]=mel(k, fs);
[fs,mediana,Longitud,media,maximo,minimo,desviacion_estandar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,amplitud]=caracteristicas(k,fs);
[medianamfcc,Longitudmfcc,mediamfcc,maximomfcc,minimomfcc,desviacion_estandarmfcc,Varianzarmfcc,Curtosismfcc,Asimetriamfcc,Vpmfcc,Shmfcc,Capmfcc]=caracteristicas_Mel(MFCCs);
muestreo(k,fs,frecuencia_acotada,amplitud);
process;
Proceso_mel;
%%
    case 3 ,
[y,fs]=grabadora;
nom=guardar(y,fs);
[y,fs]=cargar(y,nom);
L=longitud(y);
[x3,ti]=normalizar(y,fs);
[senal_filtrada,filtb,filta]=pasabajos_butterworth(x3,fs);
k=silencio(senal_filtrada,fs,nom);
fourier(k,fs,nom);
[frecuencia_acotada,amplitud]=frecuencia_C(k,fs);
[ MFCCs, FBEs, frames ]=mel(k, fs);

```



```

[fs,mediana,Longitud,media,maximo,minimo,desviacion_est
andar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,
amplitud]=caracteristicas(k,fs);
[medianamfcc,Longitudmfcc,mediamfcc,maximomfcc,minimomf
cc,desviacion_estandarmfcc,Varianzamfcc,Curtosismfcc,As
imetriamfcc,Vpmfcc,Shmfcc,Capmfcc]=caracteristicas_Mel(
MFCCs);
muestreo(k,fs,frecuencia_acotada,amplitud);
process;
Proceso_mel;

%%
    case 4,
disp('Cerrando el programa')
close all;
pause(2);
clc
fuera=1;
end

function [y,fs]=grabadora;
clc;
disp('Grabar audio');
pause;
fs=44100;
t=2;
y=wavrecord(t*fs,fs,1); %Realiza en proceso de
grabación de 2 segundos ya que el tiempo es igau a "t".
sound(y,fs);           %Reproduce el sonido grabado.
disp('Grabación realizada');pause;

function nom=guardar(y,fs)
clc;

```

```

disp('Guardar audio');
pause;
nom=input('escriba un nombre para el audio: ','s');
wavwrite(y,fs,16,nom); %Guarda el audio grabado
anteriormente a 16 bits/muestras
display('el audio se ha guardado correctamente');pause;

```

```

function [y,fs]=cargar(y1,nom);
clc;
disp('Cargar audio');
pause;
name=nom;
[y,fs]=wavread(name); %Carga los datos del audio
grabado y almacenado con cierto nombre.
disp('Audio cargado correctamente');pause;

```

```

function L=longitud(y)
fil=y.'; % Organiza los datos en vector fila.
L=length(fil); % Longitud de la señal.

```

```

function [x3,ti]=normalizar(y,fs);
%% NORMALIZACION DE LA SEÑAL "x2 es la señal
normalizada"
clc;
disp('Normalización de la señal');
pause;
L=longitud(y); %Halla la longitud de la señal.
Ts=1/fs; % Tiempo de muestreo.
ti=(1:L)*Ts; % Vector de tiempo.
x=y;
mwad=max(x); %Halla el valor máximo de la señal.
mwid=min(x); %Halla el valor mínimo de la señal.

```

```

x3=x/(max(mwad,mwid)); % se divide el vector para el
máximo y mínimo valor
plot (ti,x3)
sound(x3,fs)
guardar(x3,fs)
disp('Normalización realizada');
pause;

function f=fourier(y,fs,nom);
disp('Análisis de la señal en frecuencia');
pause;
[y,fs]=wavread(nom);
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiente potencia de 2 de la
long de y
Y = fft(y,NFFT)/L; % FFT de la señal
f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1)) ./4e-10);
amplitud=max(FDB)
title('Espectro unilateral de y(t)')
xlabel('Frecuencia (Hz)')
ylabel('dB')
xlim([0,5000])
plot(f,FDB)%%%%%%%%%%%%%%
xlim([0,10000])
w=find(FDB==amplitud);%encuentra la posición de
mw=max(w); %halla el valor más alto
freq=f(mw)%halla la frecuencia central haciendo una
relación con la amplitud pico del espectro de
frecuencia
disp('Análisis de la señal en frecuencia realizado');
pause;

```

```

function
[fs,mediana,Longitud,media,maximo,minimo,desviacion_est
andar,Varianza,Curtosis,Asimetria,Vp,Sh,Cap,frecuencia,
amplitud]=caracteristicas(y,fs)
%[y,fs]=wavread('USUARIO');
disp('Caracteristicas de la voz');
pause;
sound(y,fs)
Fs=fs;
%% parametros
mediana=median(y)%calcula la mediana
Longitud=length(y); %Longitud de y
media=mean(y)%calcula la media
maximo=max(y)%calcula el valor más alto
minimo=min(y)%calcula el valor más bajo
desviacion_estandar=std(y)%calcula la desviación
estándar
Varianza = var(y)
Curtosis = kurtosis(y) % cálculo de curtosis
Asimetria=skewness(y) % Cálculo la asimetría
%http://probestunalmlz1.wikispaces.com/cap_1_p_11
% y= número total de muestras del vector
for i=1:(length(y)-1)
    var1=abs(y(i+1)-y(i));
end
V1=(1/((length(y))-1))*var1;
for i=1:(length(y)-1)
    var2=abs(y(i));
end
V2=(1/((length(y))-1))*var2;
Vp=V1/V2; %perturbacion de la amplitud

for j=1:(length(y)-1)

```

```

        va1=abs(max(y(j+1))-max(y(j)));
end
V3=(1/(length(y)-1))*va1;
for j=1:(length(y)-1)
    va2=max(abs(y(j)));
end
V4=(1/(length(y)-1))*va2;
Sh=V3/V4 %cálculo de Shimer
%% -----
for k=1:(length(y)-1)
    v1=(y(k+1)-y(k))^2;
end
V5=sqrt((1/(length(y))))*v1;
for i=1:(length(y)-1)
    v2=abs(y(k));
end
V6=(1/(length(y)-1))*v2;
Cap=V5/V6 %perturbacion de la amplitud
f=linspace(-Fs/2,Fs/2,length(y)); %Vector creado para
el eje de frecuencias
sa=fftshift(abs(fft(y))); %transformada rápida de
Fourier a la cual se le saca el valor absoluto
filal=sa.' % Organiza los datos en vector fila.
amp=max(sa); % Pico máximo
subplot(211)
plot(f,fftshift(abs(fft(y))))%reorganiza las salidas de
FFT, moviendo el componente de frecuencia cero hasta el
centro de la matriz.
%Es útil para la visualización de una transformada de
Fourier con el componente de frecuencia cero en el
centro del espectro.
w=find(sa==amp);%encuentra la posición de
mw=max(w); %halla el valor más alto

```

```

freq=f(mw)%halla la frecuencia central haciendo una
relación con la amplitud pico del espectro de
frecuencia
disp('Cálculo de frecuencia realizado');
%%
fs=Fs;
L=length(y); % Longitud de la señal
NFFT = 2^nextpow2(L); % Siguiete potencia de 2 de la
Long de y
Y = fft(y,NFFT)/L; % FFT de la señal
f = fs/2*linspace(0,1,NFFT/2+1); % Rango de frecuencia
FDB=10*log10(2*abs(Y(1:NFFT/2+1))./4e-10);
amplitud=max(FDB);
w=find(FDB==amplitud);%encuentra la posición de
mw=max(w); %halla el valor más alto
frecuencia=f(mw)%halla la frecuencia central haciendo
una relación con la amplitud pico del espectro de
frecuencia

nam=input('Ingrese un nombre para almacenar las
características: ','s');
save
(nam,'fs','mediana','Longitud','media','maximo','minimo
','desviacion_estandar','Varianza','Curtosis','Asimetri
a','Vp','Sh','Cap','frecuencia','amplitud');
disp('Parámetros estadísticos obtenidos correctamente')
pause;

function
[n_bits,Fs,senal_muestreada,index,a,muestras]=muestreo(
y,fs)
%% -----Lectura de Datos-----
clc;

```

```

disp('Muestreo de la señal analizada');
pause;
n_bits1=input('ingrese el número de bits: ','s');
Fs1=input('ingrese el número de muestras: ','s');
n_bits=str2double(n_bits1);
Fs=str2double(Fs1);
disp('Procesando...');
[a,il]=size(y);
tiempo=a/fs;
Fs=round(Fs*tiempo);
figure(3)
plot(y);
wavplay(y,fs);
%% -----Muestreo-----
muestras=round(linspace(1,a,Fs));
senal_muestreada=zeros(1,Fs);
index=1;
for i=muestras
    senal_muestreada(index)=y(i);
    index=index+1;
end
stem(senal_muestreada);
disp('Muestreo realizado');
pause;

function
[flujo,tam,n_bits,niveles,a,muestras,vmin,vmax,fs,s_cua
ntizada,frecuencia_acotada,amplitud]=cuatizacion(y,n_bi
ts,index,senal_muestreada,a,muestras,fs,frecuencia_acot
ada,amplitud)
%% -----Cuantización-----
disp('Cuantización de la señal analizada');
pause;

```

```

vmax=max(y);
vmin=min(y);
lsb=(vmax-vmin)/(2^n_bits);
niveles=vmin:lsb:vmax;
particion=(vmin-(0.5*lsb)):lsb:(vmax+(0.5*lsb));
[il,indice_niv]=size(niveles);
% genera valores cuantizados de muestras
s_cuantizada=zeros(1,index-1);
%=====
for i=1:index-1
    for j=1:indice_niv
% si el valor es <vmin entonces será igual a vmin
        if (senal_muestreada(i)<particion(1))
            s_cuantizada(i)=vmin;
            val_binairo=dec2bin(0,n_bits);

            rango_val=val_binario(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end

        if (senal_muestreada(i)>particion(end))
            s_cuantizada(i)=vmax;
            val_binairo=dec2bin(2^n_bits)-1,n_bits;

            rango_val=val_binario(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end

        if (senal_muestreada(i)>=particion(j))

```



```

        if (senal_muestreada(i)<particion(j+1))
            s_cuantizada(i)=niveles(j);
            %Al mismo tiempo la generación de
flujo binario para cada
            %valor cuantificado
            val_binairo=dec2bin(j-1,n_bits);

            rango_val=val_binairo(:);
            rango_index=((i-1)*n_bits)+1;
            flujo(rango_index:(rango_index+n_bits-
1),1)=rango_val;
        end
    end
end
end
[tam,il]=size(flujo);
figure(4)
stairs(s_cuantizada)
ylim([vmin vmax])
set(gca,'YTick',vmin:lsb:vmax)
disp('Cuantización realizada');
pause;
codifica(flujo,tam,n_bits,niveles,a,muestras,vmin,vmax,
fs,s_cuantizada,frecuencia_acotada,amplitud)

function
cod=codifica(flujo,tam,n_bits,niveles,a,muestras,vmin,v
max,fs,s_cuantizada,frecuencia_acotada,amplitud);
%% -----Codificador-----
% de acuerdo a la codificación binaria natural
val_cod=zeros(tam,1);
for i=1:tam
    val_cod(i,1)=str2double(flujo(i,1));

```

```
end
figure(5)
stairs(val_cod');
axis([1 1000 -3 4]);
set(gca,'YTick',-3:1:4);
set(gca,'YTicklabel',{' ',' ',' ','0','1',' ',' ',' '
'});
disp('Codificación realizada');
pause;
```