



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

**TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y
COMPUTACIÓN**

**Creación de una base de datos emocional multimodal de personas
cognitivamente normales y personas con daño cognitivo.**

TRABAJO DE TITULACIÓN.

AUTOR: Valdivieso Alarcón, Daniel Rafael

DIRECTOR: Cordero Zambrano Jorge Marcos, Mgtr.

LOJA-ECUADOR

2018



Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

2018

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Magister.

Jorge Marcos Cordero Zambrano

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de fin de titulación: **Creación de una base de datos emocional multimodal de personas cognitivamente normales y personas con daño cognitivo**, realizado por Valdivieso Alarcón Daniel Rafael, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, marzo del 2018

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Valdivieso Alarcón Daniel Rafael declaro ser autor del presente trabajo de titulación: Creación de una base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo, de la Titulación de Sistemas Informáticos y Computación, siendo Jorge Marcos Cordero Zambrano director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f.

Autor Valdivieso Alarcón Daniel Rafael

Cédula 1104937402

DEDICATORIA

Este proyecto va dedicado en primer lugar a Dios quien me ha sabido dar la sabiduría para concluir de la mejor manera esta etapa universitaria en mi vida.

A mi madre Elizabeth Alarcón por ser una madre ejemplar que ha sabido guiar con sabiduría cada uno de mis pasos, a mi padre Franklin Valdivieso por los consejos y apoyo brindado,

A mi hermano Franklin Vicente Valdivieso Alarcón es mi apoyo para seguir adelante.

A mi novia Anyela por su apoyo incondicional en esta etapa de mi carrera universitaria.

A los docentes por brindarme sus conocimientos, ya que sin su aporte no hubiera sido posible culminar con éxitos mis estudios.

A mis amigos y compañeros, que siempre estuvieron presentes y con los que hemos luchado para alcanzar esta meta.

Daniel Rafael Valdivieso Alarcón

AGRADECIMIENTO

Agradezco a Dios por su amor, fidelidad y guía en este camino de mi vida.

A mis padres, hermano y familia por el apoyo y el cariño brindado.

De forma especial agradezco a mi director de proyecto de titulación Mgtr. Jorge Cordero por haberme guiado, brindarme su apoyo y dedicación para poder culminar el presente trabajo.

A mis amigos quienes a lo largo de la carrera me brindaron su apoyo y confianza para finalizar la carrera.

Daniel Rafael Valdivieso Alarcón

ÍNDICE DE CONTENIDOS

CARATULA	i
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE TABLAS	ix
ÍNDICE DE FIGURAS.....	x
RESUMEN.....	1
ABSTRACT	2
CAPITULO 1 INTRODUCCIÓN.....	3
1.1 Introducción.....	4
1.2 Planteamiento del problema y limitaciones.....	6
1.3 Justificación.....	7
1.4 Objetivos	9
1.4.1 Objetivo general.....	9
1.4.2 Objetivos específicos.	9
1.5 Estrategia o metodología de desarrollo	9
1.6 Estructura del documento.....	10
CAPITULO II.....	11
MARCO TEORICO	11
2.1 La computación afectiva.....	12
2.2 La emoción.....	13
2.2.1 Descripción de las emociones.....	14
2.2.1.1 Sorpresa.....	14
2.2.1.2 Asco.	15
2.2.1.3 Miedo.	15
2.2.1.4 Alegría.....	16
2.2.1.5 Tristeza.	17
2.2.1.6 Enfadado.....	18
2.2.2 Comparativa de la expresión vocal de las distintas emociones.	18
2.3 Sistema de reconocimiento de emociones	19
2.4 Base de datos emocionales.....	19
2.5 Definición de demencia	21
2.6 Daño cognitivo.....	21

2.7 Trabajos Relacionados.....	22
2.7.1 Base de datos deap: A database for emotion analysis using physiological signals.	22
2.7.2 Base de datos MAHNOB-HCI.	24
2.7.3 Decaf: meg-based multimodal database for decoding affective physiological responses.	25
2.7.4 Emovo corpus: una base de datos de voz emocional italiana.	27
2.7.5 Recola base de datos multimodal de colaboración remota y las interacciones afectivas.	28
2.7.6 Base de datos Rekemozio.	30
2.7.7 Comparación y puntos relevantes.	32
CAPÍTULO III METODOLOGÍA	34
3.1 Creación de la base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo	35
3.1.1 Fase 1: Diseño de la investigación.....	37
3.1.1.1 Tamaño de la muestra.....	37
3.1.1.2 Recolección.....	38
3.1.1.3 Procedimiento de obtención de datos.....	38
3.1.1.3.1 Contexto de las grabaciones.	39
3.1.1.3.2 Formatos de las grabaciones.	41
3.1.2 Transcripción.	41
3.1.2.1 Formato de transcripción.	41
3.1.2.2 Desarrollo de transcripción.	43
3.1.2.3 Resultados del proceso.	45
3.1.3 Segmentación y etiquetado del audio.	46
CAPITULO IV ANÁLISIS PARÁMETROS DE VOZ.....	48
4.1 Fase 2: análisis del audio	49
4.1.1 Análisis de los parámetros de voz.....	49
4.1.1.1 Parámetros relacionados con la prosodia.	49
4.1.1.2 Parámetros relacionados con la cualidad de la voz.	51
4.1.2 Extracción de parámetros de voz.	52
4.1.2.1 Análisis prosódico de la voz.	52
4.1.2.2 Herramientas praat y weka.	52
4.1.2.3 Rasgos prosódicos extraídos.	54
4.1.3 Técnica de análisis.	55
4.1.3.1 Métodos no supervisados.....	55
4.1.3.1.1 Clustering.....	56

4.1.3.1.2 Clustering n�merico k-means.....	57
CAPITULO V RESULTADOS	59
5.1 Evaluaci�n de resultados.	60
5.1.1 An�lisis de los resultados obtenidos mediante el m�todo de clasificaci�n cluster y el algoritmo k-means.	60
5.1.1.1 An�lisis de los resultados obtenidos por el algoritmo k-means en personas con da�o cognitivo.	62
5.1.1.1.1 Conversaci�n 1.	62
5.1.1.1.2 Conversaci�n 5.	64
5.1.1.1.3 Conversaci�n 6.	66
5.1.1.2 An�lisis de los resultados obtenidos mediante el algoritmo k-means en personas sanas.	67
5.1.1.2.1 Conversaci�n 2.	67
5.1.1.2.2 Conversaci�n 7.	69
5.1.1.2.3 Conversaci�n 10.	70
5.1.1.3 Conclusiones del an�lisis.	72
CONCLUSIONES	73
RECOMENDACIONES.....	74
BIBLIOGRAF�A.....	75
ANEXOS.....	77
Anexo A: Estructura de conversaci�n.....	78
Anexo B: Petici�n ingreso al centro del adulto mayor de Loja	79
Anexo C: Consentimiento Informado.....	80
Anexo D: An�lisis de los resultados obtenidos por el algoritmo k-means en los participantes con da�o cognitivo	81
Conversaci�n 2.....	81
Conversaci�n 3.....	82
Conversaci�n 4.....	83
Conversaci�n 7.....	84
Conversaci�n 8.....	85
Conversaci�n 9.....	86
Conversaci�n 10.....	87

ÍNDICE DE TABLAS

Tabla 1. Comparación de las características prosódicas de las emociones	18
Tabla 2. Comparación de las Bases de Datos y sus características	32
Tabla 3. Datos Generales Participantes.....	37
Tabla 4. Duración final de las grabaciones	46
Tabla 5. Emociones y características del habla	51
Tabla 6. Resultados en el reconocimiento de emociones en personas con daño cognitivo..	60
Tabla 7. Resultados en el reconocimiento de emociones en personas sanas.....	61
Tabla 8. Reconocimiento de emociones con configuración de 2 cluster	62
Tabla 9. Reconocimiento de emociones aplicando cluster 3 en la conversación 1.....	63
Tabla 10. Reconocimiento emociones aplicando 5 cluster en la conversación.....	65
Tabla 11. Reconocimiento de emociones aplicando 3 cluster	66
Tabla 12. Reconocimiento de emociones aplicando 3 cluster en la conversación 2.....	67
Tabla 13. Reconocimiento de emociones aplicando 4 cluster en la conversación 2.....	68
Tabla 14. Reconocimiento de emociones configurando 3 cluster en la conversación 7	69
Tabla 15. Reconocimiento de emociones configurando 3 cluster en la conversación 10	70
Tabla 16. Tasa de acierto en personas sanas y con daño cognitivo	72
Tabla 17. Reconocimiento de emociones configurando 3 cluster en la conversación 2	81
Tabla 18. Reconocimiento de emociones configurando 3 cluster en la conversación 3	82
Tabla 19. Reconocimiento de emociones configurando 5 cluster en la conversación 4	83
Tabla 20. Reconocimiento de emociones configurando 3 cluster en la conversación 7	84
Tabla 21. Reconocimiento de emociones configurando 4 cluster en la conversación 8	85
Tabla 22. Reconocimiento de emociones configurando 3 cluster en la conversación 9	86
Tabla 23. Reconocimiento de emociones configurando 3 cluster en la conversación 10	87

ÍNDICE DE FIGURAS

Figura 1. Enfado, sorpresa, tristeza, miedo y alegría	4
Figura 2. Ejemplo del sistema de captura de la base de datos DEAP	23
Figura 3. Ejemplo de captura de las señales para la base de datos MAHNOB	25
Figura 4. (a) Puesta en marcha experimental. (b) Un sujeto que realiza el experimento	26
Figura 5. Recopilación de datos multimodales de interacciones espontáneas en condición remota	28
Figura 6. Aplicaciones utilizadas para realizar las grabaciones.....	31
Figura 7. Proceso metodológico para realizar las grabaciones y transcripción de los videos	36
Figura 8. Entrevista a un participante en el laboratorio de inteligencia artificial de la UTPL .	39
Figura 9. Entrevista en el centro del adulto mayor de la ciudad de Loja.....	40
Figura 10. Formato del modelo de plantilla transcripción	42
Figura 11. Transcripción de un participante	45
Figura 12. Segmentación y etiquetado de la emoción en cada frase	47
Figura 13. Representación con Praat de un fichero de audio junto con el contorno de F0 ...	50
Figura 14. Representación con Praat de un fichero de audio junto con el contorno de energía	50
Figura 15. Análisis de voz praat.....	53
Figura 16. Interfaz de Weka.....	53
Figura 17. Valores numéricos correspondientes a cada parámetro de voz	55
Figura 18. Representación gráfica de un cluster	56
Figura 19. Ejemplo de clustering con k-medias.....	58
Figura 20. Distribución de resultados en personas con daño cognitivo.	61
Figura 21. Distribución de resultados en personas normales	62
Figura 22. Distribución de las emociones de tristeza y alegría mostradas en la tabla 8	63
Figura 23. Distribución de la emoción sorpresa, alegre y tristeza	64
Figura 24. Distribución de las emociones alegría, tristeza, enojo, miedo y sorpresa.....	65
Figura 25. Distribución de las emociones de alegre, sorpresa y triste en la conversación 6	66
Figura 26. Distribución de las emociones de neutro, tristeza y alegre	68
Figura 27. Distribución de las emociones de neutro, tristeza alegre y sorpresa	69
Figura 28. Distribución de las emociones neutro, sorpresa y alegría	70
Figura 29. Distribución de las emociones neutro, alegre y enfado	71
Figura 30. Distribución de las emociones tristeza, alegría y miedo	81
Figura 31. Distribución de las emociones tristeza, sorpresa y alegría	82
Figura 32. Distribución de las emociones tristeza, enfado, neutro, alegre y sorpresa	83
Figura 33. Distribución de las emociones alegre, triste y enfado.....	84

Figura 34. Distribución de las emociones enfado, alegría, sorpresa y triste	85
Figura 35. Distribución de las emociones neutro, alegre y triste	86
Figura 36. Distribución de las emociones alegre, sorpresa y tristeza	87

RESUMEN

El trabajo de fin de titulación ilustra la creación de una base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo. Luego de un meticuloso estudio de investigación bibliográfica donde se describen diferentes trabajos realizados sobre base de datos emocionales.

Se realiza un proceso metodológico para la creación de la base de datos la cual consta de tres fases: diseño de la investigación, análisis del audio por medio de parámetros de voz y por último los resultados obtenidos por el método de clasificación utilizado.

Finalmente se resaltan algunas de sus utilidades y trabajo futuro.

PALABRAS CLAVE: base de datos, emocional, multimodal, parámetros de voz, reconocimiento emocional.

ABSTRACT

The end-of-degree work illustrates the creation of a multimodal emotional database of cognitively normal people with cognitive impairment. After a meticulous bibliographic research study where, different works are described based on emotional data.

A methodological process is carried out for the creation of the database which consists of three phases: research design, analysis of the audio by means of voice parameters and finally the results obtained by the classification method used.

Finally, some of its utilities and future work are highlighted.

KEYWORDS: database, recognition emotional, multimodal, voice parameters.

CAPITULO 1
INTRODUCCIÓN

1.1 Introducción

En este primer capítulo se describe una introducción, planteamiento del problema y sus limitaciones, justificación, objetivos, metodología de este Proyecto Fin de Carrera. Asimismo, la última sección presenta la estructura de este trabajo, explicando brevemente el contenido de cada sección.

Las emociones son un elemento inherente a los seres humanos. Tanto el afecto y la emoción juegan un papel importante en nuestras vidas y están presentes casi en todo lo que realizamos. Mediante la expresión de emociones durante la interacción social se transmite información implícita importante tanto propias como estados de emociones ajenos.

Según Paul Ekman ("Ekman & Friesen, 1980) las emociones pueden ser vistas por su valor adaptativo con las tareas fundamentales de la vida. Cada emoción tiene características únicas y otras que son comunes que son producto de nuestra evolución y se distinguen así de otros fenómenos emotivos.

Para este autor, basa la existencia de las emociones básicas: enojado, miedo, tristeza, alegría, disgustado y sorpresa; en su origen evolutivo (Jose Francisco Solís Villareal, 2011).

En la Figura 1 se muestra las emociones básicas definidas por Ekman las mismas que son: enfado, sorpresa, tristeza, asco, miedo y alegría.



Figura 1. Enfado, sorpresa, tristeza, asco, miedo y alegría
Fuente:(Tobergte & Curtis, 2013)
Elaboración: Tobergte & Curtis, 2013

De tal manera que la gran mayoría de base de datos emocionales multimodales realizadas consideran dichas emociones, ya que estas son las más comunes y usadas por el hombre.

Darle la capacidad a una máquina para detectar y reconocer las emociones del ser humano ha sido un área de investigación en la cual se han hecho grandes esfuerzos en años recientes, teniendo en cuenta que entender y diferenciar las emociones expresadas por las personas es una tarea compleja aún hasta para los mismos seres humanos.(Alejandro, Valencia, Alexánder, & López, 2013)

Actualmente la mayor parte de los sistemas de Interacción Humano Computadora (IHC) incorporan sistemas tanto del habla y visión, ya que estos medios son los más naturales en la comunicación humana. Uno de los objetivos que persiguen los sistemas de IHC es que la interacción en estos escenarios sea bidireccional, para lo cual una máquina debe escuchar el mensaje del usuario y responder de manera natural.

Para alcanzar esta forma de interacción la expresión emocional debe ser reconocida y sintetizada. De esta manera, los sistemas de IHC pueden adaptarse al estado emocional del usuario, como lo hacemos los humanos al conversar, alcanzando una interacción más natural, eficiente y amigable. El reconocer el estado de ánimo de los usuarios en un sistema de IHC le brinda información relevante al sistema, retroalimentándolo y haciéndolo capaz de reaccionar y adaptarse.

La mayoría de las aplicaciones del reconocimiento de emociones, requieren que el procesamiento se realice de una forma eficiente, con buenos porcentajes de acierto y un costo computacional bajo.

Es por eso dada la importancia que se ha dado a este campo de investigación, se han construido y presentado bases de datos multimodales, las cuales brindan una importante herramienta a los investigadores que quieran enfocar sus trabajos en cuanto a las ciencias afectivas.

Estos recursos son muy adecuados para el entrenamiento de las diferentes aplicaciones emocionales, síntesis emocional.

1.2 Planteamiento del problema y limitaciones

En el aspecto tecnológico las creaciones de base de datos emocionales son muy útiles para sistemas de reconocimiento de emociones.

El área de reconocimiento automático de emociones ha sido un área de investigación muy activa en los últimos años, pero aún no hay una solución clara para este problema. Diversos inconvenientes han influido en la construcción de una solución apropiada.

Por un lado, un factor que afecta el desempeño de los reconocedores de emociones en contextos reales es la dificultad de generar bases de datos con emociones espontáneas.

Normalmente se ha trabajado con bases de datos actuadas las cuales proporcionan una inducción de emociones las cuales se ven representadas por expresiones prototípicas e intensas que facilitan la búsqueda de correlación de la característica acústica como de las expresiones faciales de las personas, sin embargo, no se han tenido buenos resultados al trasladar el conocimiento extraído de estas bases de datos a contextos reales.

Por lo cual otro reto a resolver es la identificación de un conjunto de características acústicas que permitan reconocer emociones en el habla espontánea, de tal manera, es necesario encontrar características que complementen la información que proporciona el aspecto prosódico de la voz.

Para el análisis multimodal de emociones varias bases de datos se han publicado en recientes años permitiendo así un incremento considerable en los trabajos que se realizan en este campo de investigación.

Existen diversos tipos de bases de datos emocionales cada una de ellas con características específicas como son:

- **Alcance:** características tales como el lenguaje, la descripción dada de emociones, y el número de sujetos o actores definidos.
- **Naturalidad:** características tales como si la semántica del habla se toma en cuenta o no, el material usado en la base de datos, y si mismo el texto que se ha utilizado ha logrado la expresión de diferentes emociones o no.
- **Contexto:** Modalidad del recurso es decía que sea solo de Audio o Audiovisual.

El problema fundamental del proyecto es entonces la creación de una base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo para el análisis y extracción de un conjunto de características, esta información será obtenida de las señales audiovisuales que brinden una mayor discriminación entre los diversos estados emocionales, esto a partir de las características extraídas y utilizando técnicas de aprendizaje automático.

Además, esta base de datos será para el uso de sistemas de reconocimiento de emociones y para futuras investigaciones en la enfermedad de personas con demencia como es el caso de Alzheimer.

A pesar de que los avances en el área han sido importantes, se ha comprobado que en contextos realistas aún falta mucho por hacer. Por lo tanto, es necesario proponer y realizar proyectos e investigaciones que permitan llegar a un buen desempeño del reconocimiento de emociones en aplicaciones del mundo real. Para ello, en este proyecto se propone la creación de una base de datos emocional para el uso en sistema de reconocimiento de emociones.

1.3 Justificación

Las bases de datos emocionales son muy útiles para sistemas de reconocimiento de emociones. A través de la creación de este tipo de base de datos emocional que contenga audio y video de personas cognitivamente normales y con daño cognitivo, se pretende hacer uso en sistemas de reconocimiento automático de emociones. Actualmente en la sección de Inteligencia Artificial de la “Universidad Técnica Particular de Loja” se cuenta con un sistema de reconocimiento automático “Human Emotional Recognition” que se usa con videos obtenidos del sitio web Youtube, pero no se tiene una colección de videos trabajados propia que pueda ser utilizada para diversos trabajos futuros de investigación sobre computación emocional.

Además, esta base servirá para futuras investigaciones en la enfermedad de personas con demencia como es el caso de Alzheimer.

Para el desarrollo de trabajos en el campo de la computación afectiva, se requieren conjuntos amplios de datos a partir de experimentos repetibles, por lo cual la falta de bases de datos emocional multimodales dificulta el estudio en esta área de estudio, esto puede darse por la falta de conocimiento necesaria para llevar a cabo los experimentos al igual que los costos asociados con los mismos.

Las bases de datos disponibles para el estudio de las emociones en su mayoría están construidas de manera actuada, esto es, los gestos o el discurso asociado a cada una de las emociones están representados por personas que, por indicación, gesticulan o hablan de acuerdo con la emoción que les ha sido indicada. De tal manera, no existe garantía de que el discurso o la gesticulación de esas personas sea la misma información que recogeríamos de modo espontáneo. Ya que un aspecto a considerar cuando se recogen emociones es la espontaneidad de estas.

Sin embargo, en recientes años, dada la importancia que se ha dado a este campo de investigación, se han construido y presentado bases de datos multimodales tales como Deap (Koelstra, 2012) y MAHNOB-HCI (Soleymani, Lichtenauer, Pun, & Pantic, 2012) , las cuales brindan una importante herramienta a los investigadores que quieran enfocar sus trabajos en cuanto a las ciencias afectivas.

Diferentes grupos de investigación alrededor del mundo han enfocado grandes esfuerzos en el campo de la computación afectiva, al igual que se han generado espacios en los últimos años para la publicación de trabajos específicamente en este campo, mediante congresos y revistas especializadas en la computación afectiva dada su importancia y los grandes alcances que se perciben a partir de su desarrollo.

Con el desarrollo de este proyecto, se busca apropiar el conocimiento de las ciencias del comportamiento y generar innovación en un área de estudio de gran importancia en la actualidad. Se busca entonces avanzar en el estudio y análisis de las emociones humanas con el fin de hacer uso en sistemas de reconocimientos, el desarrollo de estos, y además sea motivo de análisis para futuras investigaciones en enfermedades de personas con demencia o asociadas a la misma.

1.4 Objetivos

A continuación, se presentan los objetivos definidos para el desarrollo del presente Trabajo de Titulación.

1.4.1 Objetivo general.

Creación de una base de datos emocional multimodal de personas cognitivamente normales y personas con daño cognitivo.

1.4.2 Objetivos específicos.

- Estudiar los trabajos relacionados sobre bases de datos emocionales.
- Transcribir el audio del video con el fin de extraer el texto.
- Analizar la relevancia de los parámetros de voz.
- Aplicar técnicas de aprendizaje automático para evaluar la utilidad de los videos mediante la base datos creada, en el reconocimiento del habla emocional.

1.5 Estrategia o metodología de desarrollo

Para el desarrollo de este proyecto se emplearán varios métodos y técnicas para alcanzar los objetivos planteados.

El método de investigación empleado en este trabajo es científico experimental que como primera parte corresponde al estudio de trabajos relacionados sobre base de datos emocionales, basado en diversas técnicas de investigación que son: revisión de literatura, bases de datos y criterios de búsqueda, informe de resultados.

Además, que en su parte experimental se realizará grabaciones de videos a diferentes personas tanto cognitivamente normales como con daño cognitivo para la creación de la base de datos, todas estas personas serán de habla español.

Los videos consisten en grabar una conversación espontánea entre el entrevistado y entrevistador, con el fin de que exista una total naturalidad al momento de captura la voz e imagen de la persona.

Las grabaciones serán realizadas en un entorno sin mucho ruido para capturar el sonido de la voz sin interferencias, además un lugar donde la persona se pueda sentir cómoda.

Luego realizada las grabaciones de los videos se procederá a transcribir el audio de estos para estudiar posteriormente la influencia de los parámetros de voz.

Se va a extraer parámetros de voz que influyen en la expresión o reconocimiento facial de aquellas personas entrevistadas.

Posteriormente se va a aplicar técnicas de aprendizaje automático para evaluar el reconocimiento emocional en el habla.

1.6 Estructura del documento

En esta sección se presenta la organización de la tesis, indicando para cada uno de los cinco capítulos en los que está dividida los temas que se tratarán.

- **Capítulo 1 Introducción:** En este primer capítulo se describe una introducción, planteamiento del problema y sus limitaciones, justificación, objetivos, metodología de este Trabajo de Titulación. Asimismo, la última sección presenta la estructura de este trabajo, explicando brevemente el contenido de cada sección.
- **Capítulo 2 Marco Teórico:** En este capítulo se presentan conceptos previos acerca de que es la computación emocional, emoción, descripción de las emociones, base de datos multimodal y sus características, sistemas de reconocimiento y que es el daño cognitivo. Se presenta antecedentes de trabajos realizados de base de datos emocionales multimodales.
- **Capítulo 3 Metodología:** Se presenta en detalle cómo se realizó el proceso de la recolección y selección de los participantes, procedimiento de las grabaciones adquiridas con los participantes de la localidad, procesamiento de la información, transcripción y segmentación de la voz.
- **Capítulo 4 Parámetros de Voz:** En este capítulo se presentan los detalles y análisis de los parámetros de voz.
- **Capítulo 5 Análisis Resultados:** En este capítulo se muestra la evaluación y análisis de los archivos de audio a través del método cluster.

Para terminar, se presentan la conclusiones, recomendaciones y anexos con el desarrollo del Trabajo de Titulación.

CAPITULO II
MARCO TEORICO

En este capítulo se definen conceptos referentes a computación afectiva, emoción, descripción las emociones básicas, sistemas de reconocimiento de emociones, base de datos emocionales, daño cognitivo acerca del presente trabajo de Fin de Titulación. Se continúa presentando los trabajos relacionados acerca de base de datos emocionales. Finalmente se presenta los puntos más relevantes y una tabla comparativa sobre el estudio de las bases de datos emocionales.

2.1 La computación afectiva

En (Causa & Sosa, 2007) nos indica que la Computación Afectiva (*Affective Computing*) es una disciplina de la Inteligencia Artificial que intenta desarrollar métodos computacionales orientados a reconocer emociones humanas y generar emociones sintéticas. La fundadora de esta línea de trabajo es Rosalind Piccard, investigadora del M.I.T. (Massachussets Institute of Technology), quién publicó el libro "*Affective Computing*" en el año 2000.

Como objetivo principal dentro del área de la computación emocional o afectiva es el de capturar y procesar información emocional con el propósito de optimizar la interacción entre personas y computadoras.

La Computación Afectiva plantea dos problemáticas de las cuales se ocupa principalmente que son:

- 1- El reconocimiento de emociones (y de expresiones emotivas) humanas por parte de una computadora.
- 2- La simulación (o generación) de estados y expresiones emocionales con computadoras.

El objetivo de la primera es captar aquellos signos relacionados con la expresión de emociones y lograr interpretar estados emocionales en función de dichos signos. Este es un tema muy complejo en el que es difícil obtener precisión. De hecho, no existe una terminología universalmente consensuada a la hora de referirse a estos fenómenos.

En la segunda, se intenta que las computadoras puedan simular procesos emocionales en base a ciertos modelos. Aquí se puede reflexionar respecto a si una computadora puede realmente tener emociones, pero, esta disciplina sólo intenta simular dichos procesos de forma tal que resulten verosímiles, dejando de lado estas controversias (Causa & Sosa, 2007).

2.2 La emoción

Al hablar de Computación Afectiva y captación de emociones, es necesario definir lo que realmente es la base para cualquier tipo de investigación en esta área, que es la emoción.

La manera en que las emociones son definidas también determina el tipo de fenómenos estudiados en la investigación sobre emociones.

Cabe recalcar que, en las disciplinas dedicadas a su estudio, no hay un consenso general acerca de su especificidad, los enfoques suelen ser divergentes.

Charles Darwin (como se citó en Espinosa & García Reyes, 2010) estableció que las emociones son patrones relacionados con la supervivencia que han evolucionado para resolver ciertos problemas que una especie ha enfrentado a través de su evolución. De este modo las emociones son más o menos las mismas en todos los humanos y en particular independientes de la cultura. Aun cuando los antropólogos afirman que las emociones son productos socioculturales, varios autores han trabajado en demostrar la hipótesis de Darwin.

El psicólogo Paul Ekman (como se citó en Espinosa & García Reyes, 2010) establece seis expresiones faciales relacionadas con emociones básicas conocidas como emociones básicas.

Por un lado, se tiende a homologar el concepto de Emoción a nociones tales como “sentimiento”, “afecto”, “motivación”, “ánimo”. Por otra parte, cada abordaje atiende a diversos aspectos: algunos definen Emoción en tanto comportamientos físicos observables (llorar, reír, etc.); otros la definen en relación a cambios fisiológicos mensurables (los cuales no necesariamente son externamente discernibles); algunos piensan su definición en referencia a los tipos de experiencias conscientes involucradas en su generación; otra línea, la definirá en función de la actividad cerebral que le da soporte (Causa & Sosa, 2007).

Como se menciona (Lógica, De, Promesas, & Computación, 2015), las emociones forman parte inseparable de nuestro ser y son vitales para nuestra supervivencia y calidad de vida, relaciones con nuestro entorno y en sociedad, educación, memoria, toma de decisiones, etc. Aunque, su estudio ha sido generalmente superficial y episódico, está cambiando gracias al avance de la tecnología que ha permitido el estudio del cerebro humano.

2.2.1 Descripción de las emociones.

(Llorente, Manuel, & Martínez, 2017) en su proyecto de investigación menciona que los requisitos que debe cumplir cualquier emoción para ser considerada como básica o primaria son: tener un sustrato neural específico y distintivo, tener una expresión o configuración facial específica y distintiva, poseer sentimientos específicos y distintivos, derivar de procesos biológicos evolutivos, y manifestar propiedades motivacionales y organizativas de funciones adaptativas.

A continuación, se describe cada una de las emociones básicas estudiadas en el proyecto de (Llorente et al., 2017) y sus efectos en la expresión vocal.

2.2.1.1 Sorpresa.

La sorpresa es considerada como la emoción básica más singular. Se trata de una reacción emocional neutra, que no podemos clasificar como negativa o positiva, como agradable o desagradable. Sin embargo, la neutralidad es difícil de conseguir, así podemos tener sorpresas agradables y desagradables, pero escasas veces indiferentes.

Algunas investigaciones indican que la sorpresa tiene más en común con las emociones negativas que con las positivas (de hecho, la expresión facial de la sorpresa se confunde más a menudo con emociones negativas, especialmente miedo, que con emociones positivas).

La sorpresa se produce por lo inesperado o desconocido. Puede ser descrita como una sensación causada por algún acontecimiento repentino e inesperado. Se produce de forma súbita ante una situación novedosa o extraña y desaparece con la misma rapidez con que apareció. Se trata de la emoción más breve de todas las primarias.

Expresión vocal: Las exclamaciones de sorpresa se distinguen de las vocalizaciones de otras emociones tales como de la ira, miedo, vergüenza o placer. La emoción de sorpresa está fuertemente asociada con un tono de alto nivel. Las vocalizaciones espontáneas de la sorpresa son del tipo: ¡o!, ¡como!

Los efectos de la sorpresa sobre la frecuencia fundamental y la prosodia, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Tono medio mayor que el de la voz neutra.
- Amplio rango de F0.
- Velocidad de habla similar a la neutra.

2.2.1.2 Asco.

El asco se considera una emoción básica porque tiene un sustrato neural innato, una expresión universal también innata, un único estado motivacional-afectivo y un patrón de respuesta asociado que es relativamente estable a lo largo de distintas situaciones, culturales e incluso especies.

El asco es una sensación que se refiere en primer lugar a algo que repugna al sentido del gusto, algo percibido en ese momento o imaginado con viveza, y en segundo lugar a algo que produce una sensación parecida en el sentido del olfato, de tacto, o incluso de la vista.

Expresión vocal:

- Emisión de sonido como de aclarar la garganta.
- Sonidos guturales *aj* o *uf*.

Los efectos del asco sobre la frecuencia fundamental y la prosodia, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Tono medio bajo.
- Amplio rango de F0.
- Velocidad de locución más baja, con grandes pausas.

2.2.1.3 Miedo.

El miedo es la emoción más estudiada tanto en los animales como en el hombre, ya que es una emoción que ha despertado mucho interés.

Hay que diferenciar la emoción de miedo de la ansiedad. Mientras que el miedo hace referencia a una emoción producida por un peligro presente e inminente, encontrándose ligado al estímulo que lo genera; la ansiedad hace referencia a la anticipación de un peligro

futuro, indefinible e imprevisible, siendo la causa más vaga y menos comprensible que en el miedo.

Expresión vocal: En situaciones de miedo extremo puede existir la tendencia natural a emitir gritos de alta frecuencia o chillidos. En particular, el miedo tiende a ser asociado tanto con la elevación como con la variabilidad del tono. Cuando una persona tiene miedo se suele producir un tartamudeo asociado al “jitter segmental” que se produce en la señal de habla.

Los efectos que produce el miedo sobre los diferentes parámetros vocales, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Tono medio elevado.
- Elevada desviación de F0.
- Presenta el mayor rango de F0 de todas las emociones.
- Gran número de cambios en la curva del tono.
- Rápida velocidad de locución.

2.2.1.4 Alegría.

La alegría es el sentimiento positivo, que surge cuando la persona experimenta una atenuación en su estado de malestar, cuando consigue alguna meta u objetivo deseado, o cuando tenemos una experiencia estética (por ejemplo, la visión de un rostro agraciado o la contemplación de una bella escultura). Por lo general, la alegría es una experiencia emocional de duración breve.

Expresión vocal: Cuando una persona esta alegre tiende a elevar el tono de la voz y a aumentar su sonoridad, al tiempo que introduce un mayor número de variaciones tonales en su discurso. Cuando la alegría se manifiesta en forma de risa, se emiten una amplia variedad de sonidos, siendo los más característicos los que adoptan la forma sonora “ja-ja” o “je-je”.

Los efectos que produce la alegría sobre los diferentes parámetros vocales, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Incremento del tono medio y su rango.

- Mayor desviación de F0.
- Mayor frecuencia de las sílabas acentuadas (incremento de la velocidad de locución).
- Mayor gradiente de ascenso y descenso de F0.

2.2.1.5 Tristeza.

La tristeza es el sentimiento negativo caracterizado por un decaimiento en el estado de ánimo habitual de la persona, que se acompaña de una reducción significativa en su nivel de actividad cognitiva y conductual, y cuya experiencia subjetiva oscila entre la congoja leve y la pena intensa propia del duelo o de la depresión.

Esta emoción se plantea ante situaciones que nos suponen alguna pérdida o que nos acarrearán algún perjuicio o daño. Pero la tristeza no tiene por qué tener siempre un cariz negativo.

Expresión vocal: Cuando nos sentimos tristes, se producen descensos en la media y el rango de la frecuencia fundamental de la señal vocal, y disminuye también su intensidad. Es decir, tanto la potencia como el ritmo o tasa media con la que vibran las cuerdas vocales se atenúa sustancialmente bajo el influjo de este tipo de emoción. Nuestro tono de voz resulta más bajo y monótono, de menor sonoridad e intensidad. La fluencia verbal también se reduce, disminuyendo el número de palabras articuladas y ampliándose el tiempo necesario para su articulación. El habla se torna así cansina y lenta. Además, la tristeza puede asociarse también con cadencias tonales descendentes; esto es, el tono de voz tiende a disminuir progresivamente a lo largo de la pronunciación de la frase.

A continuación, se muestran, de forma esquemática, los efectos que produce la tristeza sobre los diferentes parámetros vocales, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Tono medio más bajo que el normal.
- Menor desviación de F0.
- Estrecho rango de F0.
- Menor frecuencia de sílabas acentuadas (velocidad de locución lenta).
- Menor gradiente de ascenso y descenso de F0.

2.2.1.6 Enfado.

El enfado es el sentimiento que surge ante una impresión desagradable, una molestia producida por alguna cosa (dicha o hecha). Se trata de un disgusto o enojo, generalmente contra otra persona.

Los parámetros de intensidad y frecuencia de la expresión vocal cambian significativamente cuando nos sentimos malhumorados. Normalmente se produce un incremento tanto en la intensidad como en la frecuencia fundamental media de la señal de voz, es decir, aumenta el vigor y el ritmo o tasa media con la que vibran las cuerdas vocales.

De esta forma, los efectos del enfado en caliente sobre los diferentes parámetros vocales, en comparación con el estado neutro, son los siguientes:

Frecuencia fundamental y prosodia:

- Tono medio alto.
- Mayor desviación de F0.
- Amplio rango de F0.
- Mayor frecuencia de sílabas acentuadas (velocidad de locución rápida).
- Mayor gradiente de ascenso y descenso de F0.

2.2.2 Comparativa de la expresión vocal de las distintas emociones.

En la tabla 1 se muestra una comparación entre los distintos valores que toman las emociones estudiadas, comparándolo con el estado neutro, para las características prosódicas: F0, rango de F0 y velocidad.

Tabla 1. Comparación de las características prosódicas de las emociones

Emoción	F0	Rango F0	Velocidad
Sorpresa	↑	↑	≈
Asco	↓	↑	↓
Miedo	↑	↑	↑
Alegría	↑	↑	↑
Tristeza	↓	↓	↓
Enfado	↑	↑	↑

Fuente: (Llorente et al., 2017)

Elaboración: (Llorente et al., 2017)

2.3 Sistema de reconocimiento de emociones

Un sistema de reconocimiento de emociones se basa en el análisis de las características particulares de las emociones para clasificar unas frente a otras. La percepción del estado anímico humano puede provenir de diferente modalidad varios canales, siendo los dos principales las expresiones faciales obtenidas a partir del canal visual y las expresiones lexico-fonéticas provenientes del habla (Ortego Resa, 2009).

Por lo tanto, un sistema de reconocimiento debe realizar un análisis multimodal sobre el estado emocional de la persona, con la información procedente de varios canales como, por ejemplo, la voz, la imagen facial, gestos y posturas, ritmo de respiración y latido del corazón, etc. Las tareas más estudiadas actualmente son el reconocimiento de emociones en el habla y en imágenes faciales.

Los sistemas de reconocimiento automático de emociones están orientados a varias aplicaciones, pero se pueden diferenciar entre dos tipos de campos de aplicaciones, entre una de ellas a mejorar la calidad de vida, y las otras que sirven para mejorar en investigaciones relacionadas con la emoción.

En la actualidad existen varias aplicaciones cuya finalidad es mejorar la calidad de vida entre las cuales tenemos servicios al cliente sensibles a la emoción, call centers, sistemas de automóviles inteligente capaces de detectar fatiga en el conductor, aplicaciones orientadas a la industria del juego y entretenimiento o sistemas de síntesis de habla emocional para discapacitados. Estos sistemas cambiarán la manera en que interaccionamos con las máquinas.

El otro grupo importante de aplicaciones está orientado a la mejora de investigaciones (por ejemplo, en psicología, psiquiatría, comportamiento humano o neurología), donde este tipo de sistemas puede mejorar la calidad de la investigación obteniendo mayor fiabilidad en las medidas y mayor velocidad en tareas manuales de procesamiento de datos sobre el comportamiento emocional (Ortego Resa, 2009).

2.4 Base de datos emocionales

Una Base de Datos Emocional (BDE) es un conjunto de archivos presentado en uno o varios formatos que contienen información emocional sobre los sujetos que han sido grabados para su elaboración. Esta información emocional se puede presentar en varios canales en forma

de expresiones faciales, voz, gestos corporales o incluso señales fisiológicas. Generalmente, los archivos de una BDE suelen estar anotados, proporcionando información útil como el tipo de emoción sentida por el sujeto, información sobre éste o el método de elicitación de emociones empleado (Tobergte & Curtis, 2013).

Las bases de datos emocionales son recursos que resultan muy adecuados para el entrenamiento de aplicaciones emocionales, tanto para sistemas de reconocimiento como de síntesis emocional, ya que en su gran mayoría suelen contener imágenes, sonidos, valores psicológicos, etc.

Además, sirven como repositorios de información para realizar futuras investigaciones sobre computación emocional.

(Tobergte & Curtis, 2013) en su proyecto presenta características a analizar en una BDE como son:

- **Formato:** es el formato en el que se presentan los archivos que contienen la información emocional. El formato puede ser vídeo, audio, imagen, texto, archivos específicos a un canal (señales fisiológicas) o cualquier combinación de éstos.
- **Modalidad:** indica qué canales de análisis emocional intervienen (e.g. expresiones faciales, lenguaje corporal, voz, señales fisiológicas... etc). Cuando una BDE contiene varios canales, se dice que es una BDE multimodal.
- **Tamaño:** número de muestras con información útil que contiene la base de datos.
- **N.º de sujetos:** en este campo se especifica el número de sujetos de los que se ha tomado información emocional para la elaboración de la base de datos.
- **Nivel de descripción:** indica el enfoque emocional en el que está anotada la base de datos, así como los distintos tipos de emociones que se evalúan en ésta.
- **Universalidad:** se dice que una base de datos emocional es universal si contiene archivos que provengan de sujetos de diferente edad, fisionomía, sexo u origen étnico.
- **Método de elicitación:** hace referencia al modo en que se extrae la información emocional de los sujetos. La elicitación se puede hacer de tres maneras: natural, donde el sujeto expresa sus emociones de manera espontánea; inducida, donde se lleva al sujeto a sentir la emoción buscada; y posado, donde el sujeto es un actor que está fingiendo la emoción.
- **Tipo de anotación:** indica la autoría de la anotación de los archivos. Las anotaciones pueden ser llevadas a cabo por anotadores expertos, por los sujetos evaluados o por ambas partes.

- **Accesibilidad:** indica la disponibilidad de la base de datos. Puede estar disponible para todos los públicos, para personal de investigación bajo pedido al autor o pudiendo ser de pago.
- **Lenguaje:** el idioma en el que está anotado la Base de Datos Emocional

2.5 Definición de demencia

La palabra demencia, etimológicamente derivada del latín demens (de: fuera, mens: espíritu, mente) indica un debilitamiento progresivo, más o menos global, de las capacidades intelectuales y afectivas, grave y definitivo (Abreu, Deterioro, & Moderado, 2015).

"Demencia es una afección ordinariamente sin fiebre y crónica, caracterizada por un debilitamiento de la sensibilidad, la inteligencia y la voluntad, con incoherencia de ideas, defecto de la espontaneidad intelectual y moral y tales son los signos de esta afección. El hombre que está en la demencia ha perdido la facultad de percibir espontáneamente los objetos, captar sus relaciones, comparar, preservar el recuerdo por completo, de lo que resulta la imposibilidad de razonar con éxito (Abreu et al., 2015).

La demencia más frecuente es la enfermedad de Alzheimer, cuyo diagnóstico clínico se realiza sobre la base de los resultados de tests psicométricos, pero el diagnóstico de certeza solo es anatomopatológico. Aunque a veces es difícil, hay que distinguir la enfermedad de Alzheimer de deterioro cognitivo propio del envejecimiento y de la pseudodemencia (Abreu et al., 2015).

2.6 Daño cognitivo

El deterioro cognitivo es un síndrome clínico caracterizado por la pérdida o el deterioro de las funciones mentales en distintos dominios conductuales y neuropsicológicos, tales como memoria, orientación, cálculo, comprensión, juicio, lenguaje, reconocimiento visual, conducta y personalidad. En muchas ocasiones el deterioro cognitivo son los primeros síntomas de enfermedades que serán diagnosticadas años más tarde, tales como Demencia o Alzheimer.

Visto como un síndrome geriátrico, es una alteración de cualquier dimensión de las funciones mentales superiores, de la cual se queja el paciente, pero se puede o no corroborar por pruebas neuropsicológicas, y que generalmente es multifactorial, tales como: depresión,

déficit auditivo y visual, hipotiroidismo, efectos adversos de medicamentos, entre otros dan lugar a una sola manifestación. Este síndrome amerita una evaluación integral para determinar si el paciente cursa con demencia, o algún otro problema que pueda ser resuelto.(General, 2012).

2.7 Trabajos Relacionados

Varias bases de datos se han publicado en recientes años permitiendo así un incremento considerable en los trabajos que se realizan en este campo de investigación. Es por ello que en el siguiente apartado se describen investigaciones de los distintos avances relevantes acerca de creación de base de datos emocionales.

2.7.1 Base de datos deap: A database for emotion analysis using physiological signals.

DEAP¹ es una base de datos diseñada específicamente para el análisis de los estados afectivos de las personas.

Se adquieren el electroencefalograma y las señales fisiológicas periféricas de 32 participantes mientras eran estimulados mediante 40 porciones de videos musicales de 1 minuto.

Los experimentos fueron llevados a cabo en 2 ambientes controlados de laboratorio con una iluminación constante. El EEG y las señales fisiológicas periféricas fueron adquiridas usando un sistema Biosemi ActiveTwo System, en un equipo de cómputo dedicado (Pentium 4, 3;2 GHz), mientras que la presentación de los segmentos de video se realiza en otro equipo dedicado (Pentium 4, 3;2 GHz) (Koelstra, 2012)(Alejandro et al., 2013).

Las señales adquiridas corresponden a los 32 canales del electroencefalograma (EEG), la respuesta galvánica de la piel (GSR), la presión sanguínea, el patrón respiratorio, la temperatura de la piel, el electrooculograma (EOG) y el electromiograma (EMG).

Adicionalmente para los 22 primeros participantes se adquirió la señal de video del rostro en una posición frontal, utilizando una video grabadora Sony DCR-HC27E (Koelstra, 2012)(Alejandro et al., 2013).

¹ <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>

Los 32 participantes estaban distribuidos equitativamente entre hombres y mujeres con edades de 19 a 37 años y cada uno realiza una valoración de cada video en términos de los niveles de activación, evaluación, gusto/disgusto, dominancia y familiaridad. Dichas valoraciones se adquieren utilizando el software Presentation de Neurobehavioral systems.

El experimento comienza con una grabación de 2 minutos de la línea basal, tiempo en el cual se le pide al paciente que se relaje para después presentar los 40 videos en 40 pruebas individuales de la siguiente forma:

- En la pantalla durante dos segundos aparece la información del número del video actual.
- Grabación de 5 segundos de línea basal.
- Presentación del video de un minuto.
- Autovaloración de los niveles de activación, evaluación, gusto y dominancia.

La base de datos está disponible para descargar en archivos pre procesados en MATLAB, donde cada archivo contiene todas las señales correspondientes a cada uno de los 40 videos de cada sujeto.

En la Figura 2 se encuentra un ejemplo del sistema utilizado para la captura de las señales al igual que el criterio de evaluación de los índices de las emociones basados en maniqués (Koelstra, 2012)(Alejandro et al., 2013).

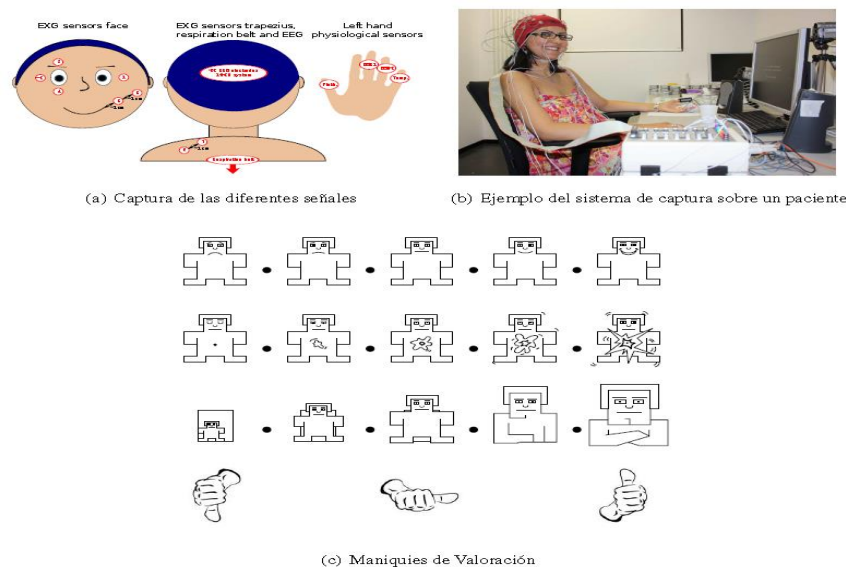


Figura 2. Ejemplo del sistema de captura de la base de datos DEAP
Fuente: <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>
Elaboración: (Alejandro et al., 2013)

2.7.2 Base de datos MAHNOB-HCI.

La base de datos MAHNOB², es una base de datos multimodal, creada dentro del campo del reconocimiento de emociones. Un setup multimodal se organiza para adquirir de forma sincronizada videos del rostro, señales de audio, seguimiento de la mirada y señales fisiológicas del sistema nervioso periférico.

El sistema utilizado para capturar las diferentes señales es el Biosemi active II system con el cual se adquieren las señales del EEG, GSR, patrón respiratorio y la temperatura de la piel. Adicionalmente la información de video es capturada desde diferentes ángulos utilizando 6 cámaras para registrar las expresiones faciales y la pose de la cabeza (Soleymani et al., 2012)(Alejandro et al., 2013).

Fueron alrededor de 30 personas de diferentes culturas que de manera voluntaria participaron en el estudio realizado, 17 mujeres y 13 hombres con edades que van desde los 19 a los 40 años. En esta base de datos el estímulo emocional se realiza de la misma forma que en la base de datos anterior, con videos que se presentan en un monitor ante los cuales los participantes deben realizar sus valoraciones en términos de los niveles de activación, evaluación, gusto/disgusto y dominancia (Soleymani et al., 2012)(Alejandro et al., 2013).

Los archivos correspondientes a los experimentos de cada sujeto, pueden ser descargados en carpetas que contienen tanto la información de video y de audio al igual que las señales fisiológicas en el formato nativo de los sistemas Biosemi ".bdf". De igual forma un archivo XML presenta la información correspondiente al experimento con las diferentes valoraciones de los participantes (Soleymani et al., 2012)(Alejandro et al., 2013).

En la Figura 3 se muestra un ejemplo del sistema de adquisición de las señales al igual que los videos de estímulo utilizados y algunas señales fisiológicas que hacen parte de las capturas realizadas.

² <http://mahnob-db.eu/hci-tagging/>

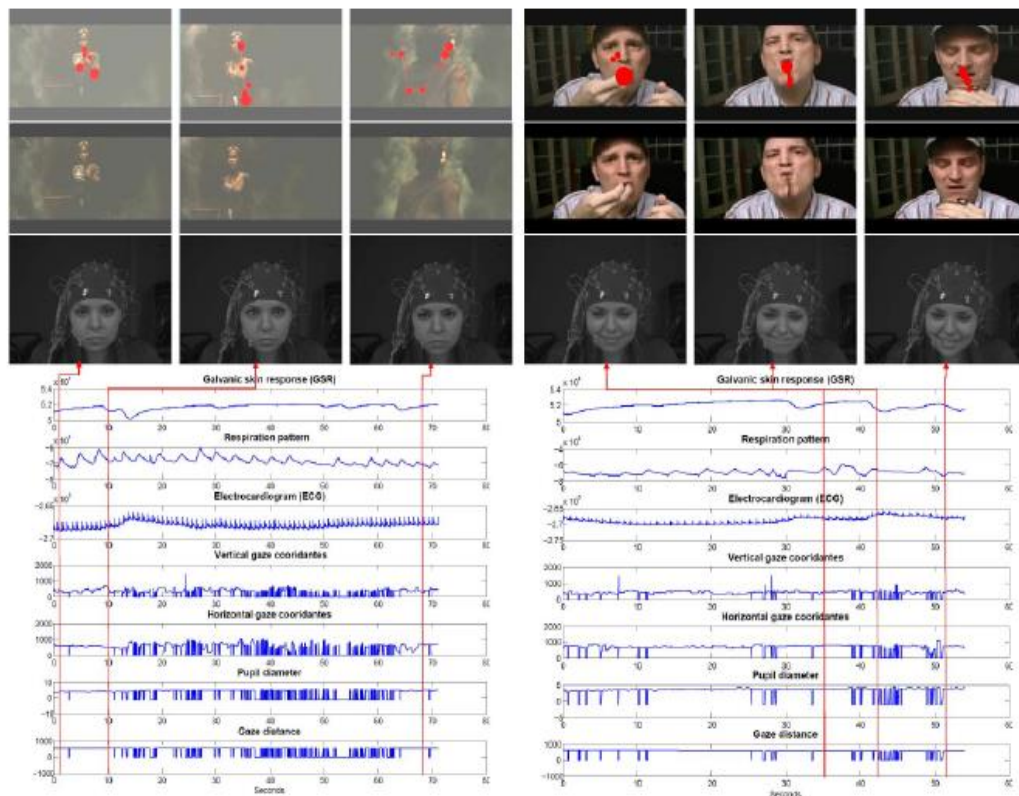


Figura 3. Ejemplo de captura de las señales para la base de datos MAHNOB
Fuente: <http://mahnob-db.eu/hci-tagging/>
Elaboración: (Soleymani et al., 2012)

2.7.3 Decaf: meg-based multimodal database for decoding affective physiological responses.

Decaf³ es una base de datos multimodal para las respuestas fisiológicas de los usuarios de decodificación afectivo contenido multimedia.

Diferente de las bases de datos tales como DEAP (Koelstra, 2012) y MAHNOB-HCI (Soleymani et al., 2012), DECAF contiene señales cerebrales adquiridas mediante el sensor de Magnetoencefalograma (MEG), que requiere poco contacto físico con el cuero cabelludo del usuario y en consecuencia, facilita la respuesta afectiva naturalista, y las respuestas emocionales explícitos e implícitos de 30 participantes a 40 segmentos de un minuto de vídeo de música utilizados en (Koelstra, 2012) y 36 clips de películas . (Abadi et al., 2015).

Todas las grabaciones se realizaron en MEG una habitación blindada con iluminación controlada.

³ <http://mhug.disi.unitn.it/wp-content/DECAF/DECAF.html>

Debido a la sensibilidad del equipo de MEG, todos los demás dispositivos utilizados para la adquisición de datos se colocaron en una habitación adyacente, y fueron controlados por el experimentador.

Se utilizaron tres ordenadores, uno para la presentación del estímulo, y otros dos para la grabación de vídeos faciales y MEG para los datos de la fisiología, como se muestra en la Figura 4, en el literal (a) encontramos el procedimiento para la captura de la información fisiológica, mientras que en el literal (b) se observa un sujeto que está realizando el experimento ya antes mencionado (Abadi et al., 2015).

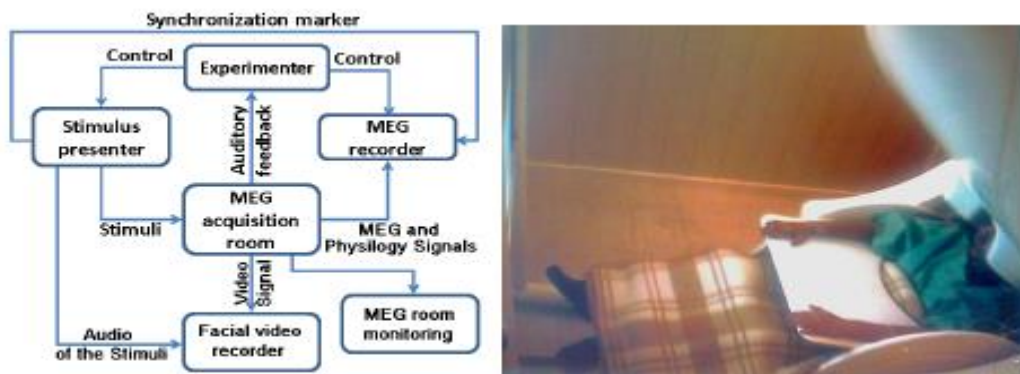


Figura 4. (a) Puesta en marcha experimental. (b) Un sujeto que realiza el experimento
Fuente: <http://mhug disi.unitn.it/wp-content/DECAF/DECAF.html> (2015)
Elaboración: (Abadi et al., 2015)

Antes del experimento a los participantes se les proporcionó un cojín, almohada y manta para relajarse durante las sesiones, además de un micrófono para informar de su estado emocional y comunicarse con los experimentadores, cada sesión, clips de música / películas se mostraban en orden aleatorio.

Tras la presentación del estímulo, el sujeto transmitió la emoción provocada en él / ella para el experimentador a través del micrófono.

Las calificaciones fueron adquiridas por:

- La excitación ('¿Qué tan intenso es su sentimiento emocional en ver el clip? ') en una escala de 0 (muy tranquilo) a 4 (muy excitado).
- Valencia ('¿Cómo se siente después de ver este vídeo? ') En una escala de -2 (muy desagradable) a 2 (muy agradable).

- Dominación en una escala de 0 (sentimiento facultada) a 4 (indefenso). (Abadi et al., 2015)

Todos los participantes fueron informados sobre el experimento y se les preguntó a dar su consentimiento informado por escrito.

2.7.4 Emovo corpus: una base de datos de voz emocional italiana.

Primera base de datos emocional, llamado Emovo⁴ aplicable a la lengua italiana. Es una base de datos construida a partir de las voces de hasta 6 actores que interpretan 14 frases que simulan 6 estados emocionales (asco, miedo, ira, alegría, sorpresa, tristeza), más el estado neutro. Las grabaciones eran realizadas con equipos profesionales en los laboratorios de la Fundación Ugo Bordoní (Costantini, Iaderola, Paoloni, & Todisco, 2014).

Se convocaron seis actores eran tres hombres y tres mujeres, que han realizado catorce (frases asertivas, interrogativas, listas) en base a seis estados emocionales básicos (asco, miedo, ira, alegría, sorpresa, tristeza) más el estado neutro.

Las actuaciones de los actores se registraron en una habitación de los laboratorios de la Fundación Ugo Bordoní en Roma. Para el experimento se contó con dos micrófonos profesionales modelo SHURE SM58LC y una grabadora digital modelo Marantz PMD670 que fueron usados. Las grabaciones se realizaron con una frecuencia de muestreo de 48 kHz, 16 bits estéreo, formato WAV (Costantini et al., 2014).

Cada sesión de grabación era aproximadamente de una hora. Después todas las frases grabadas se almacenan en la base de datos en 6 diferentes carpetas correspondientes a los 6 actores.

Todas las frases han sido registradas en una base de datos en Microsoft Access, en el que podemos ordenar los registros de acuerdo con la emoción, el texto de la sentencia, el tipo de oración, y también puede escuchar la sentencia al hacer clic en el enlace apropiado.

El número total de material grabado es de aproximadamente 10 minutos por el actor, luego de una hora para toda la base de datos. Esta base de datos EMOVO está disponible actualmente en una CD-ROM y solo serán distribuidos a través de una solicitud.

⁴ <http://voice.fub.it/activities/corpora/emovo/index.html>

2.7.5 Recola base de datos multimodal de colaboración remota y las interacciones afectivas.

Recola ⁵es un corpus multimodal espontáneo de colaboración remota e interacciones afectivas en francés.

Un total de 46 (27 mujeres, 19 hombres) personas participaron en la prueba, se registraron en grupos de dos personas dando un total de 23 equipos, que durante una videoconferencia tenían que completar una tarea que requería colaboración.

Diferentes datos multimodales como audio, video, electrocardiograma (ECG) y actividad electrodermal (EDA) se registraron continuamente y de forma sincrónica. De tal manera que 6 anotadores midieron las emociones continuamente en dos dimensiones: la excitación nominal y valencia, así como etiquetas de comportamiento social en cinco dimensiones.

Cada equipo fue recibido en una habitación dividida en dos secciones para el experimento, en el departamento de psicología del Université de Fribourg-Universität Freiburg, Suiza, el mismo contaba con cortinas gruesas que se mantuvieron cerradas durante el experimento para así reducir la variabilidad de luz solar para el video de grabación.

En la figura 5 se muestra la configuración del experimento que se utilizan en el corpus recola para recopilar datos multimodales de interacciones espontáneas en condición remota.

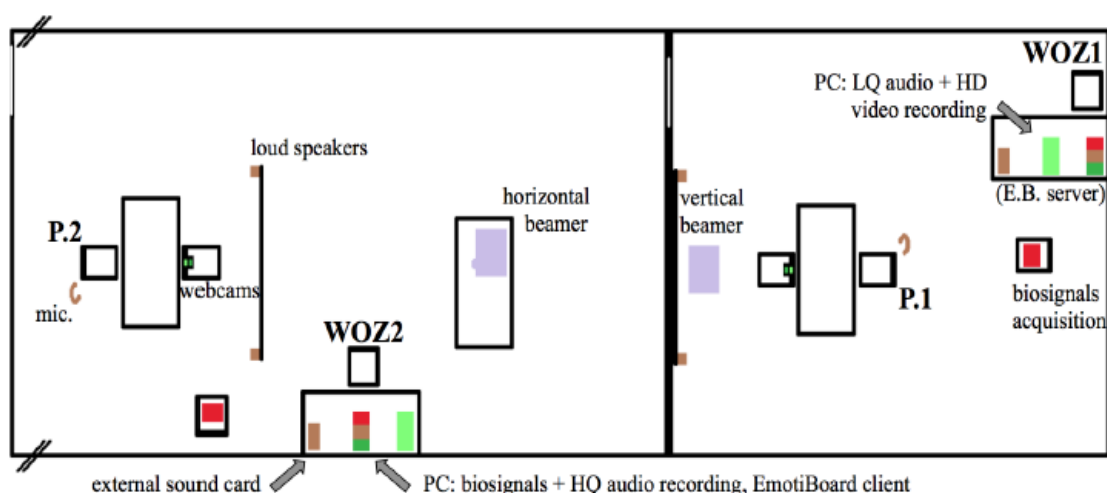


Figura 5. Datos multimodales de interacciones espontáneas en condición remota
Elaboración: (Ringeval, Sonderegger, Sauer, & Lalanne, 2013)
Fuente: (Ringeval et al., 2013)

⁵ <https://diuf.unifr.ch/diva/recola/index.html>

Estas dos secciones eran aisladas de ruidos externos, proporcionando así una adecuada ubicación para la grabación de datos.

Una vez que los participantes de un equipo se introdujeron entre sí, a continuación, que fueron separados en las dos secciones de la habitación recibieron una introducción en el experimento, se les dijo que eran participando en un estudio centrado en la comunicación entre las personas mediante el uso de herramientas asistidas por ordenador, para una duración total de aproximadamente una hora de duración.

Los participantes comenzaron a resolver de forma individual una tarea de supervivencia en caso de un desastre, con una duración máxima de 10 minutos.

Se elige este tipo de tarea por parte de los experimentadores ya que es con frecuencia utilizado en la psicología social para provocar procesos de toma de decisiones en pequeños grupos: la tarea de supervivencia. Esta tarea fue diseñada originalmente por la Administración Nacional de Aeronáutica y del Espacio (NASA) para entrenar a los astronautas antes del primer alunizaje (Ringeval et al., 2013).

Por lo cual la tarea utilizada para recoger las interacciones de colaboración espontáneo tenía que ser lo más simple posible, asegurando al mismo tiempo que las personas serían ambas impulsadas y suficientemente involucradas con respecto a sus emociones durante su comunicación con su compañero de equipo remoto.

Se promueve pidiendo a los participantes para llegar a un consenso sobre cómo sobrevivir en un escenario de desastre, como un avión estrellarse en una desierta y el área hostil. El grupo tiene que clasificar un número de elementos (por lo general 15) de acuerdo con su importancia para miembros de la tripulación para sobrevivir.

Los participantes realizaron por primera vez esta tarea de forma individual, de tal manera discutir su solución con su compañero de equipo hasta que alcanzan un consenso; la duración media era de unos 15 minutos.

Este tipo de tarea proporciona una serie de ventajas para el uso en el contexto actual. Permite modelar el proceso típico de alcanzar consenso en una compleja tarea de toma de decisiones, donde la solución al problema no era sencilla y puede requerir un uso intensivo discusión de grupo.

Al final del experimento, los participantes completaron un formulario que incluye varios autoinformes sobre la tarea, así como su consentimiento para el uso, el intercambio o la publicación de sus datos.

Los datos de audio fueron capturados por los micrófonos unidireccionales auriculares (AKG C520L) y registrado con el software Audacity a 44.1 kHz, 16bits. Una tarjeta de sonido externa (Lexicon Omega) se utilizó para dividir los datos de audio para ser utilizada simultáneamente por Skype y Audacia. Dos cámaras HD (Logitech C270) se utilizaron para cada participante. La primera cámara web sólo se capturaron los datos de vídeo y la señal se dividió en dos usando software Splitcam: uno para el Skype videoconferencia y el otro para la pantalla de visualización. Para los datos fisiológicos, se utilizó el Biopac unidad de MP36 y el software Biopac Student Lab para registrar tantas señales de EDA y ECG a 1kHz. (Ringeval et al., 2013).

2.7.6 Base de datos Rekemozio.

La base de datos Rekemozio fue creado con el objetivo de servir como un repositorio de información para llevar a cabo la investigación sobre las emociones del usuario. Esta base de datos Rekemozio se basa en los datos adquiridos a través de la interacción del usuario y los metadatos utilizados para describir y etiquetar cada interacción y proporcionar acceso a los datos almacenados y la facultad de realizar transacciones a través de ellos, por lo que la nueva información se puede agregar a la base de datos mediante el análisis de los datos incluidos en eso.

Cuando se construye la base de datos Rekemozio, el objetivo fue la adición de información descriptiva sobre las grabaciones realizadas, por lo que los procesos, tales como la extracción de parámetros de voz y funciones de vídeo, se pueden realizar actualmente en ellos.

La base de datos Rekemozio se compone de grabaciones de audio y vídeo, varios de estos realizadas por actores profesionales, mientras que otros son por aficionados. Todas estas grabaciones son en las lenguas vasca y castellano. La Figura 6 muestra las diferentes aplicaciones que se utilizan para realizar las grabaciones.

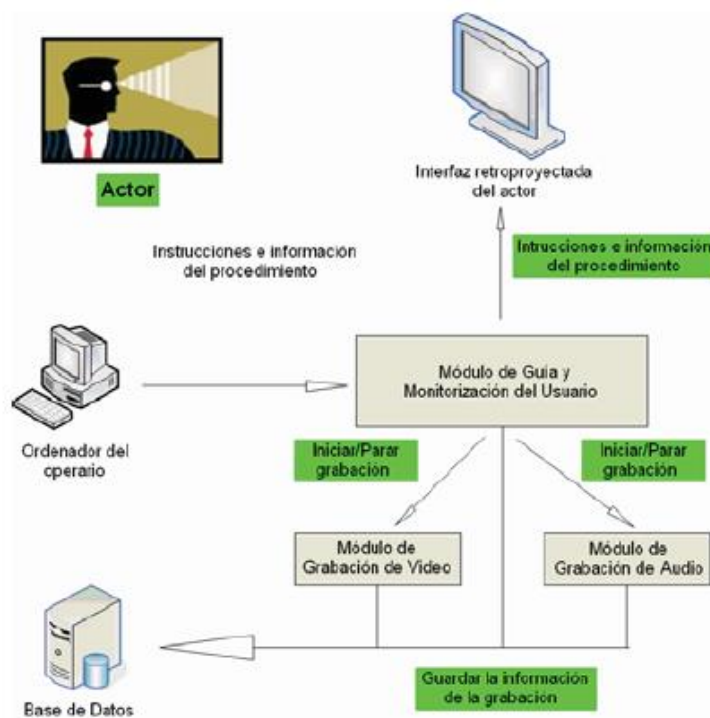


Figura 6. Aplicaciones utilizadas para realizar las grabaciones
 Fuente: (López, Cearreta, Garay, & ..., 2006)
 Elaboración: (López et al., 2006)

Se utilizaron siete emociones: las seis emociones básicas descritas por ("Ekman & Friesen, 1980 Facial-Sign-Of-Emotional-Experience.pdf," n.d.), que es la tristeza, el miedo, la alegría, la ira, sorpresa y disgusto, y una emoción neutra, lo que sugiere que estas emociones son universales para todas las culturas. Esto es interesante, teniendo en cuenta el bilingüismo de la base de datos Rekemozio.

En la base de datos existen 88 textos usados diferentes entre palabras, frases y párrafos con 154 grabaciones sobre ellos para cada actor. En un ámbito de 17 actores, siete que fueron registrados con el Idioma Vasco, mientras que 10 se registraron con el Castellano. La longitud total de las grabaciones de audio fue de 130'41" para el vasco y 166'17" para el Castellano.

Una validación de estudio normativo se llevó a cabo por los sujetos experimentales con el fin de obtener los valores afectivos de cada grabación y ver lo que la validez del material grabado y los valores afectivos para cada grabación. Los resultados obtenidos muestran que el material grabado en la base de datos RekEmozio fue identificado correctamente por 57 sujetos experimentales, con una precisión media del 66,5% para las grabaciones de audio (Álvarez, Sierra, Arruti, Lópezgil, & Garay-Vitoria, 2015).

2.7.7 Comparación y puntos relevantes.

En la Tabla 2 se presenta el resumen de los trabajos más relevantes en el ámbito de la creación de base de datos emocionales multimodales.

Algunos puntos relevantes de estos trabajos son los siguientes:

- En su totalidad de trabajos realizados han sido de forma inducida para capturar las emociones mostradas por los participantes, como son DEAP (Koelstra, 2012) Y DECAF (Abadi et al., 2015) que a través de video clips de música y película provocan estímulos de emociones en los participantes.
- En general las bases de datos se encuentran en idiomas extranjeros como son Inglés (Koelstra, 2012) (Abadi et al., 2015) (Soleymani et al., 2012), Vasco (Álvarez et al., 2015), Francés (Ringeval et al., 2013) e Italiano (Costantini et al., 2014), por lo cual con este trabajo se pretende tener una base de datos local de habla español.
- Se aprecia una mayor cantidad de trabajos realizados que han considerado un alto número de participantes para la creación de las bases de datos emocionales como son 32 participantes (Koelstra, 2012) , 27 participantes (Soleymani et al., 2012), 42 participantes (Abadi et al., 2015), 46 participantes. (Ringeval et al., 2013)

Tabla 2. Comparación de las Bases de Datos y sus características

Trabajo	Base de Datos	Numero Participantes	Actuada o Espontánea	Inducida o Natural	Audio	Video	Idioma
(Koelstra, 2012)	DEAP	32	Espontánea	Inducida	No	Si	Inglés
(Soleymani et al., 2012)	MAHNOB-HCI	27	Espontánea	Inducida	Si	Si	Varios Idiomas.
(Abadi et al., 2015)	DECAF	42	Espontánea	Inducida	No	Si	Clips de Película en Inglés.
(Costantini et al., 2014)	EMOVO	6	Actuada	Inducida	Si	Si	Italiano.
(Ringeval et al., 2013)	RECOLA	46	Espontánea	Inducida	Si	Si	Francés.
(Álvarez et al., 2015)	RekEmozio	17	Actuada	Inducida	Si	Si	Vasco. Castellano.

Fuente: El Autor
Elaboración: El autor

Estos puntos representan áreas de investigación que pueden extenderse con técnicas, recursos, análisis, condiciones y alternativas de desarrollo adicionales. El presente trabajo puede contribuir en este aspecto mediante las siguientes propuestas:

- Desarrollar una base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo con personas de la ciudad de Loja-Ecuador.
- Establecer los parámetros de una base de datos emocional para el desarrollo y análisis de desempeño de sistemas de reconocimiento multimodal.
- Una base de datos que esté disponible al público para futuras investigaciones de habla español.
- Analizar la relevancia de los parámetros de la voz que influyen en la expresión y el reconocimiento emocionales, y a través de aplicación de diferentes técnicas de aprendizaje automático poder evaluar sus utilidades en el reconocimiento del habla emocional

Para el desarrollo del presente trabajo, se ha tomado características del procedimiento usado de la base de datos RekEmozio (Álvarez et al., 2015), ya que usa un enfoque multimodal y cuestiones principales a considerar como: ámbito, naturalidad, contexto, descriptores y accesibilidad que se deben de tener en cuenta para describir una base de datos.

La base de datos que se pretende crear se realizara a través de grabaciones de audio-video, pero con personas no profesionales o actores que se les induzca las emociones si no que lo hagan de forma natural por medio de una conversación, este escenario es relevante ya que se transmite en un contexto espontaneo de interacción.

El escenario del comportamiento espontáneo garantiza interacciones emocionalmente de forma natural, ya las conductas mostradas surgen de un contexto real en el que el conjunto de señales verbales y no verbales producidos es libre e ilimitada.

Pero este escenario de interacción espontánea es difícil de lograr, ya que incluye varias cuestiones éticas, como que las personas discutan acerca de cosas privadas, y saber que se está realizando una grabación.

CAPÍTULO III

METODOLOGÍA

La mayoría de las bases de datos que existen actualmente están desarrolladas en otros idiomas Inglés (Koelstra, 2012) (Soleymani et al., 2012) (Abadi et al., 2015), Vasco (Álvarez et al., 2015), Italiano (Costantini et al., 2014) y que son grabaciones realizadas con actores, por tal razón se pretende la creación de una base de datos emocional multimodal para el idioma Español, en condiciones ambientales de iluminación y poco ruido para la captura de los datos con el fin de contar con datos más reales.

El propósito de este trabajo es además incluir información descriptiva sobre las grabaciones para permitir la identificación y categorización de cada una de ellas.

3.1 Creación de la base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo

Se recopila un nuevo conjunto de datos compuestos de audio-videos cuya característica principal es de naturaleza espontánea, las grabaciones fueron realizadas a personas con diferentes patologías en lo que se refiere a daño cognitivo como es el Alzheimer, demencia entre otras y personas que se encuentren sanas. Estas grabaciones fueron relajadas y utilizando métodos no invasivos.

Los tiempos de grabación son más cortos para el grupo de personas con daño cognitivo esto se deben al hecho de que los pacientes encuentran en el hablar más de esfuerzo que las personas sanas: hablan más lentamente, con pausas más largas y con más tiempo dedicado a buscar la palabra correcta.

El esquema detallado del proceso metodológico de las grabaciones de audio-video y transcripción a texto de estos se puede ver en la figura 7.

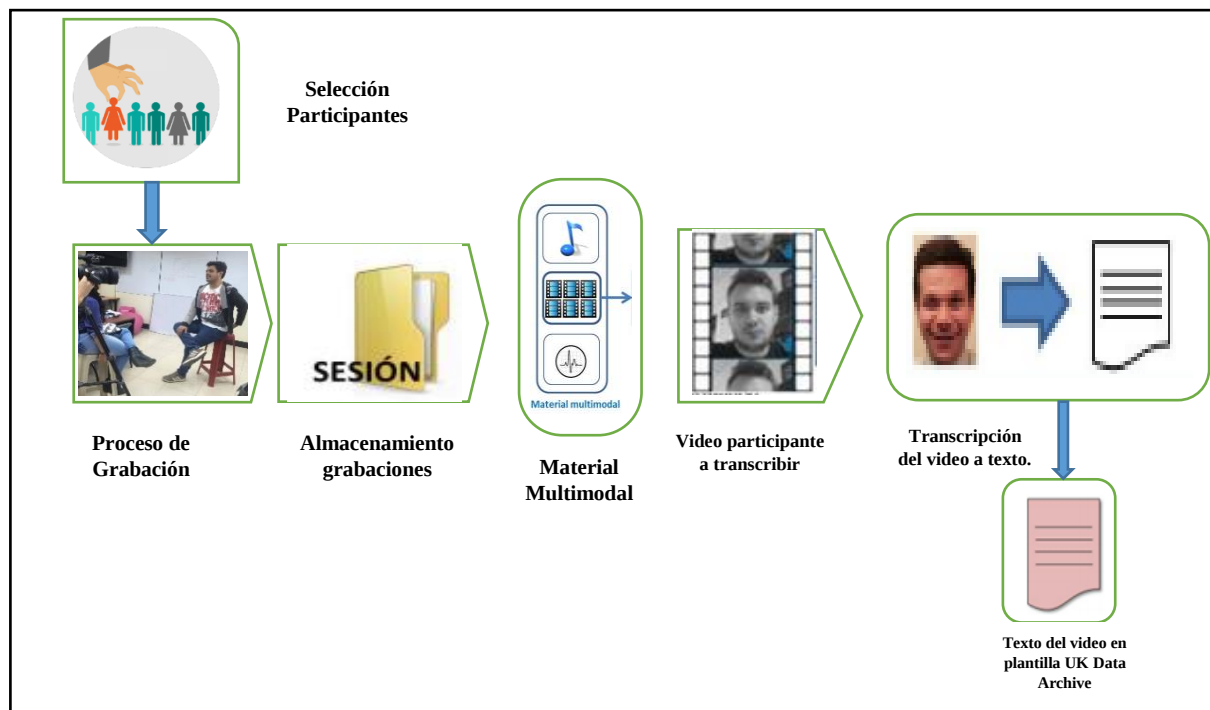


Figura 7. Proceso metodológico para realizar las grabaciones y transcripción de los videos

Fuente: El Autor

Elaboración: El Autor

Como se muestra en la figura 7, en primera instancia se procede a realizar la selección de personas con problemas cognitivos y sanos, luego de aquello existe un tiempo de grabacion en donde se realizan una entrevista personal en donde se efectuan preguntas de carácter personal e intrafamiliar para conocer un poco más de los participantes, a continuacion se almacena la información en varios formatos para su posterior analisis. Posteriormente se arroja el material del proceso anterior con los resultados y entonces se da paso a la transcripcion del mismo a texto para tener un respaldo material fisico del mismo. Finalmente se pasa el diseño en formato texto del video plantilla Uk Data Archive.

A continuación, se propone los criterios generales para el diseño y especificación de la base de datos de pacientes con daño cognitivo y personas cognitivamente normales que fue grabada con el objeto de posibilitar el trabajo desarrollado en este proyecto, y su relación con la búsqueda de parámetros para la detección automática emociones mediante el análisis del habla.

3.1.1 Fase 1: Diseño de la investigación.

El diseño de la investigación se enfoca a seguir una estructura ordenada para conseguir el objetivo planteado, conociendo el tamaño de la muestra, la recolección de la información, la obtención de los datos y el análisis de los resultados.

A continuación se aborda cada uno de los temas mencionados en el diseño de la investigación.

3.1.1.1 Tamaño de la muestra.

En el presente Trabajo de Titulación se considera una muestra de 20 personas para la creación de la base de datos emocional multimodal, de las cuales 10 personas cognitivamente normales y 10 personas con daño cognitivo, fueron realizadas las grabaciones a través de una conversación espontánea entre el entrevistado y el entrevistador.

Todos los participantes son pertenecientes a la ciudad de Loja para realizar conversaciones espontáneas y proceder así mismo a la grabación. Estos participantes no fueron actores profesionales.

Los participantes para la base de datos emocional estuvieron dentro del grupo de edades para personas cognitivamente normales entre los 25 a 40 años y aquellas personas con daño cognitivo entre los 65 a los 85 años. Para tener una pronunciación estándar todas las personas son del habla español.

Finalmente, un total de 10 personas normales y 10 personas con algún tipo de daño cognitivo fueron considerados para la base de datos emocional. Sus datos generales se presentan en la Tabla 3.

Tabla 3. Datos Generales Participantes

Participantes	Idioma	N° Participantes	Género (H-M)	Edad Promedio	Emociones
Sanos	Español	10	4/6	32.5	Categórico: 6 emociones básicas
Daño Cognitivo		10	5/5	80	

Fuente: El Autor
Elaboración: El Autor

3.1.1.2 Recolección.

La recolección de datos se puede considerar como un procedimiento conveniente para la obtención de información necesaria para el diseño de la investigación. Por lo cual según la naturaleza de la investigación la técnica utilizada es la entrevista.

Según (Gomez Bermeo, 2017) (Rodríguez, 2011) la define como: “La entrevista es una técnica personal que permite la recolección de la información en profundidad donde el informante expresa o comparte oralmente y por medio de una relación interpersonal con el investigador su saber (opiniones, creencias, sentimientos, puntos de vista y actitudes) respecto a un tema o hecho.”

La entrevista tuvo una duración promedio de cinco minutos donde el entrevistador utiliza un cuestionario como instrumento de investigación, que se encuentra compuesto por preguntas objetivas acerca de su información personal con el fin de obtener respuestas verbales e identificar las emociones a las interrogantes planteadas por parte del entrevistado.

3.1.1.3 Procedimiento de obtención de datos.

En la entrevista se hablan sobre temas relacionados a datos personales, familiares y actividades que desarrollan en su vida. Por lo cual se elabora una conversación o entrevista estructurada para los dos grupos de participantes (ver Anexo A).

A través de esta conversación se pretende obtener información para la base de datos a crear, y de esta manera puedan expresar emociones básicas como: Enfadado, Felicidad, Neutro y Tristeza, de una manera natural y espontánea a preguntas relacionadas con su vida diaria.

En el caso de las personas con daño cognitivo se tiene que realizar una petición de autorización al Patronato de Amparo Social de la ciudad de Loja (ver Anexo B), para el ingreso al centro del Adulto Mayor, ya que en este lugar se toma un número significativo de la muestra total de personas que sufren algún tipo de enfermedad asociada al daño cognitivo como es Alzheimer, Demencia entre otras.

De igual forma se tiene que pedir el consentimiento informado a los representantes de los familiares que se encuentran en este centro (ver Anexo C) para la realización de las grabaciones, así como el uso de los datos en este proyecto.

3.1.1.3.1 Contexto de las grabaciones.

Se hizo uso del laboratorio de Inteligencia Artificial de la “Universidad Técnica Particular de Loja” para grabar los recursos de la base de datos con aquellas personas sanas.

En la parte frontal del participante se sitúa una cámara para las grabaciones de video, se tiene que tener una perspectiva de todo el rostro de la persona.

Al comenzar las grabaciones se debe realizar un proceso de calibración de las cámaras para así comprobar que tanto el video como el audio se recibían de manera adecuada.

Como se tiene que realizar una conversación espontánea, para que las personas entrevistadas puedan demostrar sus expresiones y emociones con total naturalidad, se pidió colaboración amigos, familiares o personas cercanas al entrevistador para que pueda haber total confianza al momento de realizar las conversaciones.

En la figura 8 se muestra el momento en el que se realiza la conversación con un participante del grupo control, en el laboratorio de Inteligencia Artificial de la “Universidad Técnica Particular de Loja”.



Figura 8. Entrevista a un participante en el laboratorio de inteligencia artificial de la UTPL
Fuente: El Autor
Elaboración: El Autor

Se tiene la colaboración de una persona para el manejo de la cámara para realizar las grabaciones, mientras se mantenía la conversación entre las dos personas. Se tuvo especial cuidado en mitigar todas las posibles fuentes de distorsión de sonido para que no afecten al proceso de grabar audio.

Las grabaciones con aquellas personas con daño cognitivo, en este caso como se ha mencionado en el apartado 3.1.1.3 se realiza con personas adultas mayores de edad comprendidas entre 65 a 85 años, en el Centro del Adulto Mayor de la Ciudad de Loja.

Al realizar una conversación con estas personas, en este caso se pide la autorización para el ingreso a este centro (ver Anexo A). Una vez aceptada la autorización de ingreso al centro, se solicita por parte de la coordinadora encargada de este lugar que antes de proceder a las grabaciones, se tenga un periodo de tiempo para compartir por parte de la persona que realiza la entrevista con aquellas personas adultos mayores. De la misma forma se informa y pide el consentimiento (ver Anexo B) a los representantes de los familiares de cada una de las personas a tener la conversación y así mismo a grabar.

Ya obtenido el consentimiento por parte de los representantes de cada una de las personas que se encuentran en este centro, se procede a realizar las grabaciones. Las grabaciones se las realiza en el patio del centro ya que cuenta con un espacio amplio, con iluminación y sin ruidos externos que puedan interferir al momento de la captación de los datos en la cámara.

En la figura 9 se muestra la entrevista que se realizó a una participante en los patios del centro del Adulto mayor de la ciudad de Loja.



Figura 9. Entrevista en el centro del adulto mayor de la ciudad de Loja
Fuente: El Autor
Elaboración: El Autor

A las 20 personas que se les realiza las grabaciones previamente se les solicita sus datos demográficos para determinar su influencia al analizar el habla o características faciales, tales como edad, idioma, etc. Además, a través de un consentimiento informado se tuvo autorización para que el uso de los datos obtenidos es netamente con fines académicos para la creación de la base de datos y confidenciales.

3.1.1.3.2 Formatos de las grabaciones.

Las conversaciones fueron captadas por una cámara de video HD (Panasonic AVCCAM) modelo AG-AC160A, el tipo de archivo de los videos es de extensión. MTS es un formato de video de alta definición y con una frecuencia de muestreo de 48,000 Hz.

3.1.2 Transcripción.

La transcripción es una traducción entre formas de datos, con mayor frecuencia para convertir grabaciones de audio a texto en la investigación cualitativa. Debe coincidir con los objetivos analíticos y metodológicos de la investigación. La transcripción es parte del proceso de análisis, y también de mejora en el intercambio y la reutilización potencial de los datos de una investigación cualitativa (Data-archive.ac.uk, 2016).

Los datos cualitativos recogidos de grabaciones audiovisuales son óptimos para realizar la transcripción de estos archivos a texto.

Para tener una base de datos emocional multimodal más completa, se realiza la transcripción de las grabaciones de audio-video de los participantes con el propósito de tener un respaldo de la información en digital, también sirva como estudio en investigaciones de base de datos de voz, donde se relaciona en el tiempo el estímulo textual con las muestras de voz.

3.1.2.1 Formato de transcripción.

(Margarita, 2016) en su investigación realiza las transcripciones de audio a texto considerando el formato de plantilla (Data-archive.ac.uk, 2016) para asegurar la calidad y consistencia de los datos. Esta plantilla es estándar y cuenta con varias directrices en lo que se refiere al estilo de las transcripciones.

Cada una de las transcripciones debe contener los siguientes parámetros:

- Identificador único, un nombre o número.
- Diseño coherente en toda una colección de datos o proyecto de investigación.
- Cabecera de documento u hoja de cubierta con la entrevista o detalles de eventos, tales como la fecha, lugar, nombre del entrevistador y entrevistado.
- Utilizar las etiquetas de los altavoces para indicar la secuencia de pregunta / respuesta o la toma de turnos en una conversación.
- Tener saltos de línea entre su vez se lo lleva.
- Utilizar seudónimos para anonimizar la información de identificación personal.

En la Figura 10 se presenta el formato utilizado para el desarrollo de las transcripciones.

Study Name:	Interview number:	
Depositor:	Interview ID:	
Interviewer:	Date of interview:	
<u>Information about interviewee</u>		
Date of birth:	Marital status:	
Gender:	Occupation:	
Geographic region:		
Y=Interviewee Entrevistado		
I=Interviewer (Entrevistador)		
 The work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England and Wales Licence. To view a copy of this licence, visit creativecommons.org/licenses/by-nc-sa/2.0/uk.		

Figura 10. Formato del modelo de plantilla transcripción
Fuente: (Data-archive.ac.uk, 2016)
Elaboración: El Autor

3.1.2.2 Desarrollo de transcripción.

Todos los videos se transcriben manualmente para extraer frases habladas, así como la hora de inicio y finalización de cada enunciado oral. La transcripción se realiza utilizando sólo el audio sin información visual. Cada vídeo contiene de 20 a 30 emisiones. La transcripción es un proceso muy fiable, pero a su vez toma mucho tiempo.

A continuación, en la Figura 11 se presenta la transcripción de la entrevista realizada a una persona que tiene daño cognitivo, en la plantilla se encuentran algunos datos demográficos del participante.

Study Name: Personas con Alzheimer Interview number: 1

Depositor: Interview ID: Angélica Salvador



Interviewer: Daniel. Date of interview:

Information about interviewee

Date of birth: Marital status: Soltera

Gender: Femenino Occupation:

Geographic region: Loja

Y=Interviewee Entrevistado

I=Interviewer (Entrevistador)

I: Hola a la María, tiene que saludar al Guillermo a su tía, a sus sobrinos,

Y: Sobrinos; sorpresa, 00:16 a 00:17

I: Y a mí, no me saluda

Y: Como que para que tanta cosa; alegre, 00:21 a 00:23

I: Vamos hacer una novela que se va a llamar Alicia el País de las Maravillas. Como se llama usted Srta. ?

Y: Alicia Angélica Salvador Pérez; alegre, 00:36 a 00:43

I: Cuantos años tiene

Y: Creo q voy a cumplir 40; alegre, 00:45 a 00:47

I: 40 que jovencita y de donde es?

Y: De Guayaquil; alegre, 00:52 a 00:53

I: Hay de Guayaquil, por eso es tan bonita

Y: Hay muchas gracias; alegre, 00:54 a 00:56

I: Y sus papás

Y: Mi papitos eran; triste, 00:57 a 00:58

Y: El unito 40 mismo no me acuerdo; miedo, 01:02 a 01:04

I=Interviewer (Entrevistador)

Y: y mi mamita tampoco me acuerdo; triste, 01:05 a 01:10

I: Y cuántos hermanos tiene?

Y: Yo tengo Guillermo, Cesar, Edmundo tres; neutro, 01:25 a 01:28

I: Y novios?

Y: Novios no tengo; alegre, 01:32 a 01:33

I: Porque? si eres bonita;

Y: Me da miedo casarse; miedo, 01:39 a 01:41

I: Y que es de su tía

Y: Mi tía; neutro, 01:44 a 01:46

Y: como mi madre Rosarito; triste, 01:49 a 01:55

I: Y donde está su tía

Y: en la casa; sorpresa, 01:55 a 01:56

I: ah en su casa esta

Y: en la casa de ella; sorpresa, 01:58 a 01:59

I: Como fue su infancia?

Y: Mi infancia fue muy buena con mis papás; triste, 02:11 a 02:12

I: Y su época de colegio

Y: también porque en la Pérez Bazaes me hicieron educar; neutro, 02:13 a 02:20

I: En Quito o en Guayaquil

Y: En quito; neutro, 02:22 a 02:23

I: Y como es su vida actual, cuénteme, se siente feliz, triste

Y: A ratos triste; triste, 02:30 a 02:32

I: Porque?

Y: Otros ratos feliz; triste, 02:32 a 02:33

I: Porque triste?

Y: Siempre hay inconvenientes no; neutro, 02:35 a 02:39

Y: entonces no voy estar riéndome; neutro, 02:43 a 02:48

I: Y porque estas feliz a veces?

Y: Alguna cosa buena se presenta; miedo, 02:56 a 02:59

Y: entonces uno cambia y se deja de ser tristonaa; sorpresa, 03:03 a 03:06

I: Y cómo se lleva con sus compañeras del centro

Y: Todas son muy buenas; miedo, 03:18 a 03:20

I: Y se quieren mucho todas?

Y: Claro; alegre, 03:23 a 03:24

I: Y que extrañas de tu vida?

Y: Nada; triste, 03:30 a 03:31

I: Entonces en otras palabras si te sientes bien aquí?

Y: Si, claro; miedo, 03:37 a 03:39

I: Le quiere mandar un saludo a su hermano Guillermo? A ver mándele un saludo a Guillermito

Y: No, que venga a verme; alegre, 03:55 a 03:57

Y: Y ahora qué fecha?; sorpresa, 04:02 a 04:05

I: 5 de Mayo del 2015

Y: Y ya mismo viene el loco; sorpresa, 04:16 a 04:17

I: Cual Loco?

Y: Mi hermano es mayo que mí, loco; alegre, 04:19 a 04:25



The work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England and Wales Licence. To view a copy of this licence, visit creativecommons.org/licenses/by-nc-sa/2.0/uk.

Figura 11. Transcripción de un participante
Fuente: El Autor
Elaboración: El Autor

3.1.2.3 Resultados del proceso.

Luego de completar el proceso de grabación, el resultado obtenido es una colección de audio-video de los 20 participantes, de los cuales 10 corresponden a personas normales y los otros 10 a personas con daño cognitivo.

Las grabaciones en audio y video están dadas para el idioma español. Por otro lado, se guarda información descriptiva sobre las grabaciones como se puede ver en la Tabla 3 (formatos de grabación y de transcripción).

Tabla 4. Duración final de las grabaciones

Participantes	Idioma	Duración audio/video	Formato
Personas Sanas	Español	70'45"	wav/mp3
Daño Cognitivo		60'27	

Fuente: El Autor

Elaboración: El Autor

Para las grabaciones del audio de los videos, se archiva las transcripciones en tipo texto con el tipo de plantillas (Data-archive.ac.uk, 2016) .

3.1.3 Segmentación y etiquetado del audio.

En un segundo paso los diálogos fueron segmentados en enunciados o frases, donde la señal de audio se almaceno por separado para cada frase.

A este proceso también se le denomina parametrización, donde un archivo de audio se divide la señal acústica en una serie de segmentos que posteriormente son analizados de manera individual.

(Manuel & Sánchez, n.d.) se refiere a un corpus de habla expresiva con emociones espontáneas denominado base de datos "FAU AIBO" donde las grabaciones fueron segmentadas automáticamente en turnos usando una pausa de 1 segundo. Cinco personas escucharon los enunciados en orden secuencial y anotaron cada palabra independientemente del resto como neutro (por defecto) o perteneciente a una emoción determinada. En este corpus la mayoría de los enunciados son solamente órdenes cortas y la información se etiqueta a nivel de palabra o frase. Para otorgar una clase a una palabra se hizo una votación entre las opiniones de los etiquetadores. Si tres o más coinciden se atribuye la etiqueta emocional a la palabra.

Como en el trabajo realizado en (Gomez Bermeo, 2017) se realiza una etiquetación manual a nivel categórico de la emoción a la frase de los participantes basado en su expresión facial y su voz.

En el presente trabajo cada archivo de audio ha sido segmentado a nivel de enunciados o frases, luego se realiza un etiquetado manual de forma subjetiva de las emociones descritas en la sección 2.2.1, que son percibidas en los segmentos de audio, proceso que se repite hasta lograr observar correctamente la emoción por parte del observador.

En la Figura 12 se muestra la segmentación a nivel de frase y el etiquetado de la emoción que se realizó a una persona con daño cognitivo por parte del anotador.

A	B	C	D
Sra. Olimpia			
	Frases	Etiqueta	Tiempo
1	Si María Olimpia soy?	alegre	01:24 a 01:27
2	García Córdova	alegre	01:30 a 01:35
3	Cuántos dijo? 91	sorpresa	01:37 a 01:40
4	Viuda, de unos 30 años ya	alegre	01:43 a 01:44
5	No, una vecita nomás	alegre	01:50a 01:53
6	Tuve 7, dos muertos y 5 vivos	alegre	01:57 a 02:01
7	Entre los vivos son cuatro mujeres y un varón	alegre	02:02 a 02:06
8	yo sabía decir cuatrocientas y un bataplan	sorpresa	02:07 a 02:11
	Mi vida de niña feliz, porque mis papá y mi mamá me		
9	querían arto	alegre	02:18 a 02:20
10	más que a ninguno de todos	alegre	02:21 a 02:25
11	éramos 9 hermanos éramos	triste	02:26 a 02:31
12	pero yo era la más mimada	alegre	02:32 a 02:34
13	Hasta ahora, pero ya murieron mis papás	alegre	02:35 a 02:39
14	Todas me miman	alegre	02:45 a 02:46
15	entonces yo era la más mimada yo era la sexta	alegre	02:52 a 02:54
16	por eso mi hermana la mayor como una muñeca me tenía	alegre	02:55 a 03:04
	y dice esta le vine comprando a la Olimpita hagámole		
17	rapidito un vestido.	alegre	03:05 a 03:13
18	mi mami era un señora que vendía telas así	alegre	03:14 a 03:27
	esta esta buena para la Olimpita llévele y así ella me		
19	mimaban	triste	03:28 a 03:33
20	Yo de Otavalo	sorpresa	03:35 a 03:37

Figura 12. Segmentación y etiquetado de la emoción en cada frase
Fuente: El Autor
Elaboración: El Autor

El resultado de esta segmentación se puede utilizar para la extracción de parámetros acústicos segmentales. Estos parámetros pueden ser tanto de prosodia como de calidad de voz, y utilizados en aplicaciones donde se requiera un modelado de estos. Ejemplos de uso son el reconocimiento de emociones (análisis) y la síntesis del habla (análisis y síntesis).

CAPITULO IV
ANÁLISIS PARÁMETROS DE VOZ

4.1 Fase 2: análisis del audio

El análisis de los audios de los videos se realiza de forma manual, la gran cantidad de información que se puede obtener de la voz ya que existe evidencia de que un amplio rango de características lingüísticas tiene significado emocional. Estas características pueden ser acústicas o prosódicas. Por lo cual se necesita describir de modo adecuado los estados emocionales que aparecen en el discurso.

4.1.1 Análisis de los parámetros de voz.

Una de las preguntas más importantes para el reconocimiento automático de las emociones en el habla es qué características se deben extraer de la señal de voz. Estudios previos muestran que es difícil encontrar características de voz específicas que puedan usarse como indicadores confiables de la emoción presente en el habla. (Álvarez et al., 2007)

Para mejorar los resultados de trabajos anteriores, se han calculado nuevos parámetros o características auditivas que se pueden extraer de un segmento de audio, pero sólo algunas son más descriptivas al momento de querer identificar una emoción. Entre los rasgos que describen mejor una emoción están el Pitch o F0, energía, entre otros.

4.1.1.1 Parámetros relacionados con la prosodia.

La prosodia caracteriza los rasgos suprasegmentales percibidos en el habla y por lo tanto se refiere a sílabas, palabras, frases o locuciones enteras en lugar de únicamente a fonemas.

(Santiago Planet Garcia, Jose Antonio Morán Moreno, 2006) en su trabajo explica que la prosodia se refiere a la percepción acústica de los sonidos, a los que define por su tono, sonido y duración. El tono es la impresión perceptiva que produce en el oyente la F0 de la onda sonora; el sonido es la fuerza percibida en los sonidos influida tanto por la intensidad de la onda sonora como por la F0, las características espectrales y la duración del sonido; y la duración es la longitud percibida de un sonido, influenciada por cambios en el sonido.

A pesar de que estas características perceptivas no tienen equivalentes únicos en la señal de voz existen equivalentes acústicos que se adaptan muy adecuadamente como la F0, que se asemeja al tono percibido, y la energía de la señal, que se asemeja al sonido. Se agrupan en tres categorías según se relacionen con la F0, la energía o la duración:

- **Parámetros relacionados con la F0:** Pueden calcularse en todos los segmentos en los que se ha fragmentado el archivo de voz, aunque realmente tiene sentido en aquellas zonas que presentan una periodicidad clara, es decir, en los sonidos sonoros como, por ejemplo, las vocales. Se puede medir en Hertz (Hz).

En la figura 13 se muestra el fichero de audio de un participante junto con el contorno de F0 de color azul, que es una línea regular continua y sostenida.

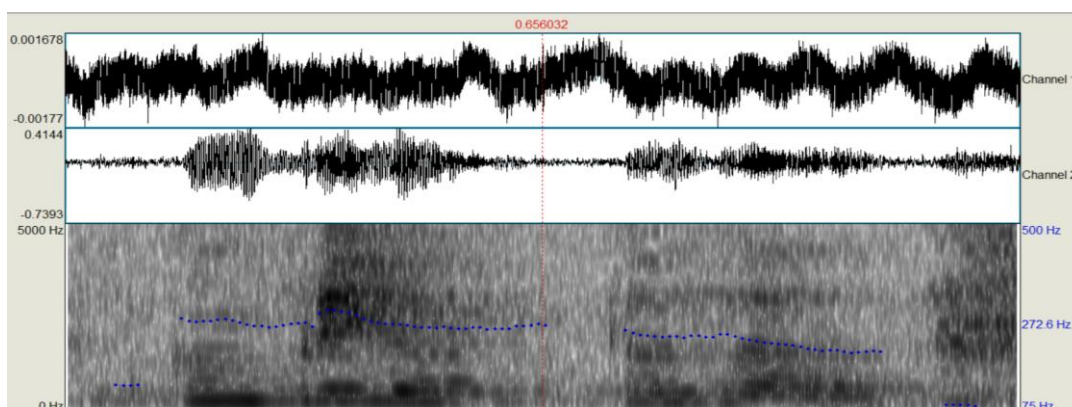


Figura 13. Representación con Praat de un fichero de audio junto con el contorno de F0

Fuente: El Autor

Elaboración: El Autor

- **Parámetros relacionados con la energía:** Como su propio nombre indica, miden la energía del fragmento acústico analizado. Se puede medir en decibelios (dB).

En la figura 14 se muestra el fichero de audio junto con el contorno de energía de color amarillo.

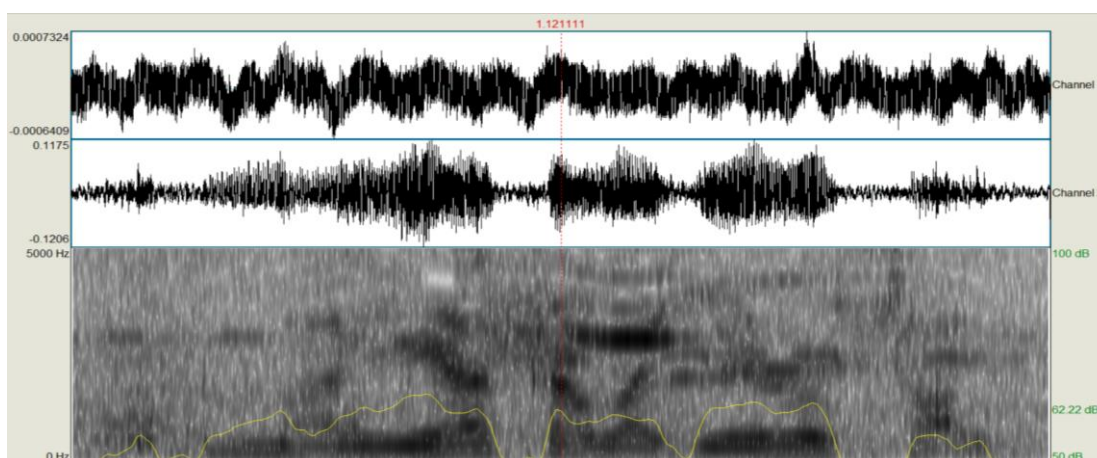


Figura 14. Representación con Praat de un fichero de audio junto con el contorno de energía

Fuente: El Autor

Elaboración: El Autor

4.1.1.2 Parámetros relacionados con la calidad de la voz.

En el trabajo realizado en (Santiago Planet Garcia, Jose Antonio Morán Moreno, 2006) se menciona que la calidad de la voz se basa en las perturbaciones en la F0 y la energía de la señal que se representan por distintos parámetros.

Se pueden destacar los siguientes parámetros relacionados con la calidad de voz:

- **Jitter.** Mide las variaciones a corto plazo del periodo fundamental debidas a fluctuaciones en los tiempos de apertura y cierre de los pliegues vocales de un ciclo al siguiente, describiendo un ruido que aparece en forma de modulación en frecuencia.
- **Shimmer.** Mide las variaciones a corto plazo de la amplitud de la forma de onda entre ciclos, describiendo un ruido que aparece como una modulación en amplitud.

En la Tabla 5 se presenta un resumen de las relaciones entre las emociones y los parámetros del discurso. Como se observa únicamente aparecen cinco emociones estas corresponden con las emociones primarias o básicas.

Tabla 5. Emociones y características del habla

	Ira	Felicidad	Tristeza	Miedo	Disgusto
Velocidad del Habla	Ligeramente acelerada	Acelerada retardada	Pausada	Muy acelerada	Mucha más acelerada
Margen de F0	Muy amplio	Muy alto	Ligeramente más estrecho	Muy amplio	Ligeramente más amplio
Calidad de Voz	Procedente del pecho	Estridente	Resonante	Irregular	Retumbante
Intensidad	Alta	Alta	Baja	Normal	Baja
Pulso Glotal	Pendiente fuerte y alto ancho banda	Pendiente fuerte	Pendiente suave y ancho banda	Pendiente muy fuerte y gran ancho	Pendiente fuerte

Fuente: (Ortego Resa, 2009)
Elaboración: El Autor

Existe en general una relación conocida entre el habla y las emociones primarias. Las medidas del discurso que parecen ser buenas indicadoras de estas emociones son medidas acústicas continuas, tales como las relacionadas con la variación del discurso, el rango, la intensidad y la duración.

4.1.2 Extracción de parámetros de voz.

En la sección presente, se tratará automáticamente las evidencias en el habla de los participantes con daño cognitivo y normales, y describir los métodos utilizados para extraer los datos prosódicos de la base de datos.

4.1.2.1 Análisis prosódico de la voz.

Una de las áreas en la dentro de la lingüística, es el estudio de la prosodia que da información sobre los rasgos fónicos dado que influye la expresión oral, tales como la acentuación, la tonalidad o la cadencia. Dichos rasgos no son iguales en cada persona, por lo tanto, se puede distinguir los rasgos de una persona a otra. Además, existen factores tales como, la edad, nivel cultural, el estado físico y emocional de cada individuo que influyen en las características prosódicas.

Estudios modernos han presentado que personas que sufren ciertas enfermedades, como el síndrome de la apnea-hipopnea durante el sueño (o SAHS) o el Alzheimer, exhiben diferencias significativas en determinados rasgos fónicos y melódicos del habla con respecto a los de personas sanas (Abreu et al., 2015).

4.1.2.2 Herramientas praat y weka.

Para el análisis prosódico de las voces grabadas se utiliza la herramienta formada particularmente para este propósito: el programa Praat⁶, diseñado y desarrollado para el análisis fonético del habla por Paul Boersma y David Weenink, del Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam (Boersma, Paul 2001).

Esta herramienta tiene un lenguaje de programación propio, por tanto, nos hace posible escribir código según las especificaciones que nos son necesarias y ejecutarlo por medio de esta herramienta.

En la Figura 15 se muestra el análisis de voz a un segmento de audio en el programa Praat que es bastante sencilla y de fácil manejo para cualquier usuario.

⁶ www.praat.org

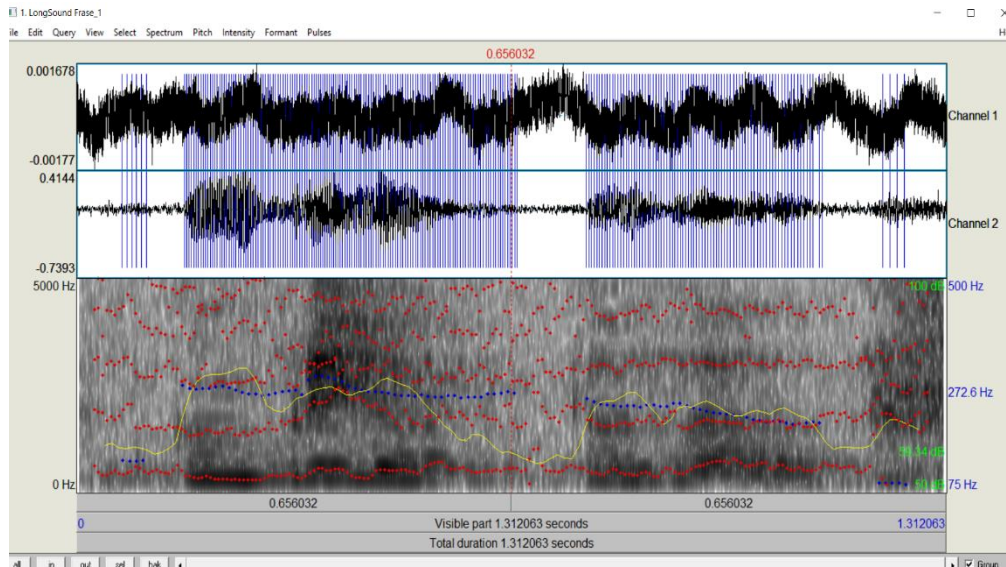


Figura 15. Análisis de voz praat
Fuente: <http://www.fon.hum.uva.nl/praat/>
Elaboración: El Autor

Weka⁷ es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o llamar desde su propio código Java. Weka contiene herramientas para el preprocesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático. (Weka,2015)

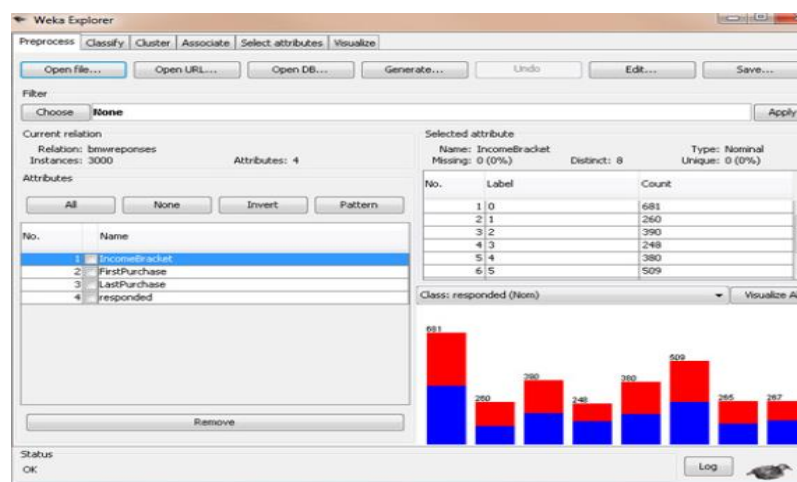


Figura 16. Interfaz de Weka
Fuente: <http://www.ibm.com/developerworks/library/os-weka2/> (2015)
Elaboración: El Autor

⁷ <https://www.cs.waikato.ac.nz/ml/weka/>

Con el uso de la herramienta Weka se puede analizar los datos recogidos a partir de la herramienta Praat como se hace referencia en la figura 16. Por lo que con este sistema Weka es muy fácil usar para llevar a cabo tareas de minería de datos y manejo en forma matricial. Este sistema arroja inmediatamente y de forma directa los datos obtenidos del proceso. Por lo tanto, en este trabajo se utilizará las librerías de Weka para extraer datos de interés a partir de los calculados por Praat y llevaremos a cabo diferentes modelos estadísticos para tratar de obtener los mejores resultados.

4.1.2.3 Rasgos prosódicos extraídos.

Como ya hemos explicado que se obtiene los parámetros prosódicos que fueron definidos en la Sección 4.1.1 por la herramienta Praat, y a partir de estos parámetros, mediante Weka se obtiene los rasgos necesarios para evaluar a través de los algoritmos de aprendizaje la base de datos creada en el habla. A continuación, se muestran los parámetros prosódicos calculados.

- **Frecuencia fundamental (F0).** Se define como el ciclo periódico de la señal de voz, siendo el resultado de la vibración de los pliegues vocales. Su medida habitual es el hercio (Hz), que da una medida de los ciclos por segundo.
- **Intensidad.** Medida de la energía de la onda acústica. Habitualmente se utiliza una transformación logarítmica de la amplitud de la señal, llamada decibelio (dB), que representa mejor la percepción humana del sonido.
- **Jitter.** Parámetro que caracteriza la perturbación de F0 debida a fluctuaciones en los tiempos de apertura y de cierre de los pliegues vocales de un ciclo al siguiente.
- **Shimmer.** Parámetro que caracteriza la perturbación en la intensidad debida a fluctuaciones en la amplitud de un ciclo al siguiente.

En la Figura 17 se muestra los valores resultantes correspondientes a cada parámetro de voz, pitch, jitter, Shimmer e intensidad en los diferentes segmentos de voz de una persona con daño cognitivo.

	ID	NOMBRE	ID_FRASE	FRASES	ETIQUETA	PITCH	JITTER	SHIMMER	INTENSIDAD
1	1	Jose Santander	v1o1	Buenos días	alegre	130	2.1	21	82
2	1	Jose Santander	v1o2	Curiosa	alegre	145	1.3	20	81
3	1	Jose Santander	v1o3	José Rafael Santander	alegre	160	3.1	19	83
4	1	Jose Santander	v1o4	No sé cuantos años tendré	alegre	141	2.8	21	75
5	1	Jose Santander	v1o5	Yo nací en el año de 1906	alegre	161	1.7	17	81
6	1	Jose Santander	v1o6	Con mi esposa	alegre	169	2.4	19	82
7	1	Jose Santander	v1o7	Clara Antonia Castellanos	alegre	163	2.8	20	82
8	1	Jose Santander	v1o8	Es mayor para mí con seis años	alegre	148	2.5	19	80
9	1	Jose Santander	v1o9	El matrimonio	neutro	156	2.5	22	79
10	1	Jose Santander	v1o10	Sentirse bien o mal	alegre	166	3.1	20	80
11	1	Jose Santander	v1o11	el matrimonio a veces bien a veces mal	sorpres	168	1.4	16	78
12	1	Jose Santander	v1o12	la mujer mas pléitista	alegre	134	1.9	20	80
13	1	Jose Santander	v1o13	que el hombre	alegre	148	2.7	19	81
14	1	Jose Santander	v1o14	que la mujer no es tan buena que se diga es pícara	sorpres	154	3.3	19	77
15	1	Jose Santander	v1o15	Porqué es pléitista	engo	154	1.9	17	81
16	1	Jose Santander	v1o16	y en veces la mujer por mejor por allá	sorpres	123	1.7	19	82
17	1	Jose Santander	v1o17	Mejor de lejos, que bonito es ver a la mujer de lejos	alegre	158	2	19	79
18	1	Jose Santander	v1o18	Porqué la mujer es más pícara que el hombre	sorpres	136	1.3	18	79
19	1	Jose Santander	v1o19	bueno bueno de repente viendo a la mujer, le simpatizo o no le simpatizo	sorpres	140	3.6	17	83
20	1	Jose Santander	v1o20	Cómo	triste	145	1.5	15	80
21	1	Jose Santander	v1o21	De muchacho entre a aprender panadería	sorpres	134	1.5	17	76
22	1	Jose Santander	v1o22	me hice panadero, con esa profesión me case	triste	135	2.6	18	75
23	1	Jose Santander	v1o23	Si	triste	126	1.6	16	77
24	1	Jose Santander	v1o24	Bueno... bien días así, días	triste	168	3.2	20	77
25	1	Jose Santander	v1o25	cinco hijos	triste	171	2.8	21	73
26	1	Jose Santander	v1o26	no sé cada cual llega el destino, se casan y se van	alegre	157	1.8	18	73
27	1	Jose Santander	v1o27	Yo con mi esposa	triste	138	1.7	14	79
28	1	Jose Santander	v1o28	con mi esposa nomás	triste	125	2.4	19	77
29	1	Jose Santander	v1o29	Buenos días, buenos días	alegre	153	2.7	19	76

Figura 17. Valores numéricos correspondientes a cada parámetro de voz

Fuente: El Autor

Elaboración: El Autor

4.1.3 Técnica de análisis.

En este apartado se abordarán los aspectos más relevantes de la técnica de aprendizaje empleadas para el reconocimiento automático de emociones en el habla a través de los audios de la base de datos creada.

4.1.3.1 Métodos no supervisados.

(Bastidas Mendieta, 2017) señala en su investigación que el objetivo del aprendizaje no supervisado es encontrar "patrones interesantes" en los datos, esto se puede llamar descubrimiento de conocimiento y es un problema mucho menos bien definido, ya que no se nos dice que tipo de patrones buscar.

El aprendizaje no supervisado, es el extremo opuesto aprendizaje supervisado, donde se deja que el algoritmo encuentre un patrón oculto en una carga de datos. Con el aprendizaje no supervisado no hay respuesta correcta o incorrecta; es solo un caso de ejecutar el algoritmo de aprendizaje automático y ver que se producen los patrones y los resultados. (Bastidas Mendieta, 2017)

En la clasificación no supervisada diremos que hemos obtenido una buena clasificación cuando los grupos creados sean homogéneos respecto a los individuos que los forman y heterogéneos entre sí.

Dada la naturaleza del problema de reconocimiento de emociones este tipo de técnicas serán muy útiles cuando las bases de datos disponibles se recojan en entornos no controlados y los individuos reflejen emociones no indicadas a priori por el investigador.

4.1.3.1.1 Clustering.

Para (Bell, 2014) (como se cito en Bastidas Mendieta, 2017), refiere que si se reducen todas las definiciones de clustering que existen se puede decir que es "la organización de un grupo de objetos que comparten características similares."

Clustering es utilizado en múltiples campos así, por ejemplo:

- Internet: Las redes sociales utiliza la agrupación para determinar las comunidades de usuarios.
- Negocios: investigación de mercado usan mucha agrupación para definir grupos de clientes.
- Cumplimiento de la ley: se ejecuta clustering y otros algoritmos de aprendizaje automático para predecir cuándo y dónde crímenes futuros sucederán.
- Informática: para agrupar los resultados de sensores. Por ejemplo, pensando en un sensor de temperatura, es posible agrupar fecha y la hora en contra de la temperatura.

En la Figura 18 se muestra que hay tres grupos distintos de los datos; cada uno de esos grupos es un cluster.

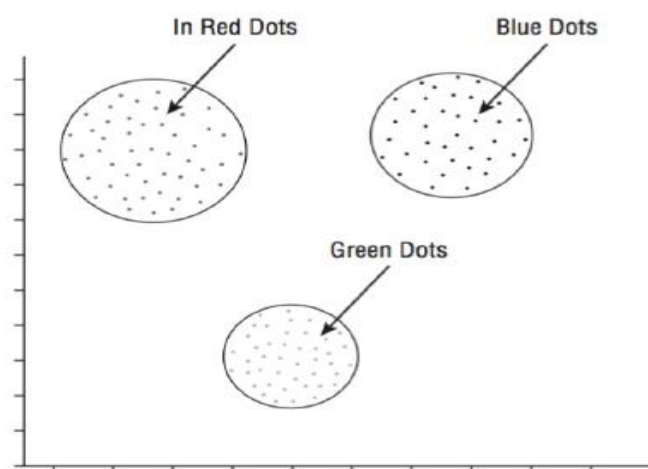


Figura 18. Representación gráfica de un cluster
Fuente: (Bastidas Mendieta, 2017)
Elaboración: (Bastidas Mendieta, 2017)

4.1.3.1.2 Clustering numérico *k*-means.

Uno de los algoritmos más utilizados para hacer clustering es el *k*-medias (kmeans), que se caracteriza por su sencillez. En primer lugar, se debe especificar por adelantado cuantos clusters se van a crear, éste es el parámetro *k*, para lo cual se seleccionan *k* elementos aleatoriamente, que representaran el centro o media de cada cluster. A continuación, cada una de las instancias es asignada al centro del cluster más cercano de acuerdo con la distancia Euclídea que le separa de él. Para cada uno de los clusters así contruidos se calcula el centroide de todas sus instancias. Estos centroides son tomados como los nuevos centros de sus respectivos clusters. Finalmente se repite el proceso completo con los nuevos centros de los clusters. La iteración continúa hasta que se repite la asignación de los mismos ejemplos a los mismos clusters, ya que los puntos centrales de los clusters se han estabilizado y permanecerán invariables después de cada iteración. (José Manuel Molina López, 2004)

Para obtener los centroides, se calcula la media según se trate de atributos numéricos o simbólicos.

A continuación, en la Figura 19, se muestra un ejemplo de clustering con el algoritmo *k*-medias. En este caso se parte de un total de nueve ejemplos o instancias, se configura el algoritmo para que obtenga 3 clusters, y se inician aleatoriamente los centroides de los clusters a un ejemplo determinado. Una vez inicializados los datos, se comienza el bucle del algoritmo.

En cada una de las gráficas inferiores se muestra un paso por el algoritmo. Cada uno de los ejemplos se representa con un tono de color diferente que indica la pertenencia del ejemplo a un cluster determinado, mientras que los centroides siguen mostrándose como círculos de mayor tamaño y sin relleno. Por último, el proceso de clustering finaliza en el paso 3, ya que en la siguiente pasada del algoritmo (realmente haría cuatro pasadas, si se configurara así) ningún ejemplo cambiaría de cluster. (José Manuel Molina López, 2004)

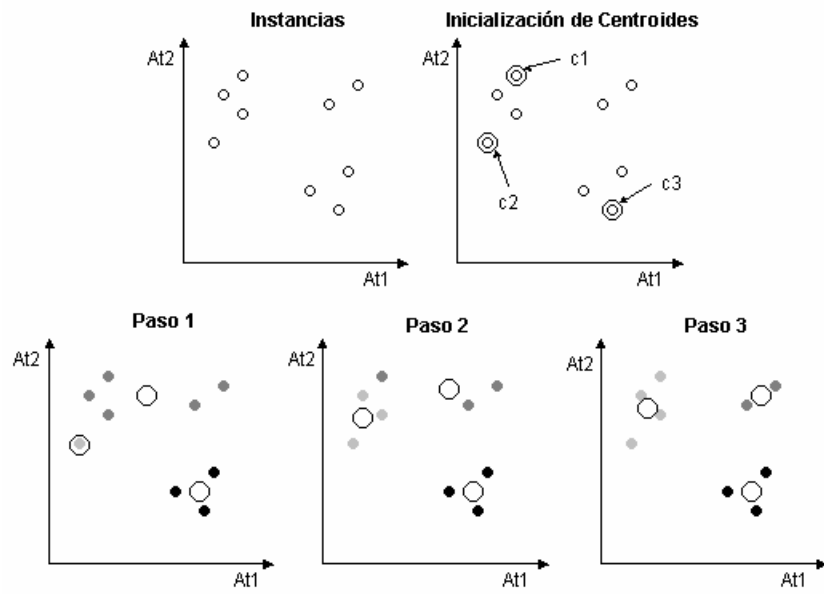


Figura 19. Ejemplo de clustering con k-medias
Fuente: (José Manuel Molina López, 2004)
Elaboración: (José Manuel Molina López, 2004)

CAPITULO V
RESULTADOS

5.1 Evaluación de resultados.

En este capítulo se presenta la evaluación y análisis de los archivos de audio de nuestra base de datos emocional obtenidos de las conversaciones de cada participante. Por medio de los diferentes parámetros de voz se puede reconocer las emociones en el habla.

El método de agrupamiento que se utiliza para el presente proyecto es cluster con el algoritmo k-means, es uno de los más utilizados y al tener un conjunto de datos se los agrupa para la correcta interpretación de los resultados.

5.1.1 Análisis de los resultados obtenidos mediante el método de clasificación cluster y el algoritmo k-means.

A continuación, se presenta un resumen y análisis de las diferentes muestras de voz de las conversaciones tanto para personas con daño cognitivo como sanas. Se ha implementa el método de agrupamiento cluster, a través del algoritmo k-means para obtener el mejor resultado en el reconocimiento emocional del habla como lo muestra la Tabla 6.

La tasa de acierto es el resultado de la comparación entre los valores cluster y el etiquetado subjetivo de las muestras de voz que indican las diferentes emociones básicas.

Tabla 6. Resultados en el reconocimiento de emociones en personas con daño cognitivo.

Conversaciones	Número de clusters	Emoción	Tasa de acierto
Conversación 1	2	Alegre-Triste	0.53
Conversación 2	3	Alegre-Triste-Miedo	0.26
Conversación 3	3	Alegre-triste-sorpresa	0.38
Conversación 4	5	Alegre-triste-enfado-sorpresa-neutro	0.40
Conversación 5	5	Alegre-triste-enfado-miedo-sorpresa	0.56
Conversación 6	3	Alegre-triste-sorpresa	0.55
Conversación 7	3	Alegre-triste-enfado	0.52
Conversación 8	4	Alegre-triste-sorpresa-enfado	0.56
Conversación 9	3	Alegre-triste-neutro	0.55
Conversación 10	3	Alegre-triste-sorpresa	0.60

Fuente: El Autor
Elaboración: El Autor

La figura 20 muestra cómo están distribuidos los valores en base a las conversaciones con personas con daño cognitivo y la tasa de acierto.

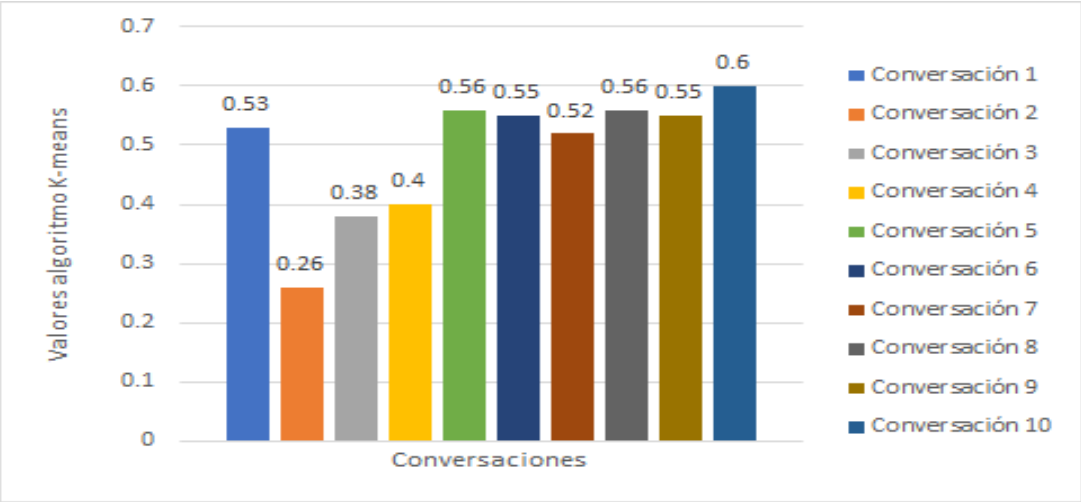


Figura 20. Distribución de resultados en personas con daño cognitivo.
Fuente: El Autor
Elaboración: El Autor

A continuación, se muestra en la Tabla 7 los resultados obtenidos aplicando el método de agrupamiento cluster en cada conversación y la tasa de acierto en personas sanas.

Tabla 7. Resultados en el reconocimiento de emociones en personas sanas

Conversaciones	Número de Clusters	Emoción	Tasa de acierto
Conversación 1	3	alegre-triste-neutro	0.55
Conversación 2	3	alegre-triste-neutro	0.55
Conversación 3	3	alegre-triste-sorpresa	0.60
Conversación 4	5	alegre-triste-enfado-neutro	0.60
Conversación 5	5	alegre-triste-enfado-neutro	0.55
Conversación 6	3	alegre-triste-sorpresa	0.54
Conversación 7	2	neutro-alegre	0.60
Conversación 8	3	alegre-sorpresa-neutro	0.56
Conversación 9	3	alegre-triste-neutro	0.60
Conversación 10	3	alegre-enfado-neutro	0.55

Fuente: El Autor
Elaboración: El Autor

La figura 21 muestra cómo están distribuidos los valores de cada conversación en personas sanas y la tasa de acierto.

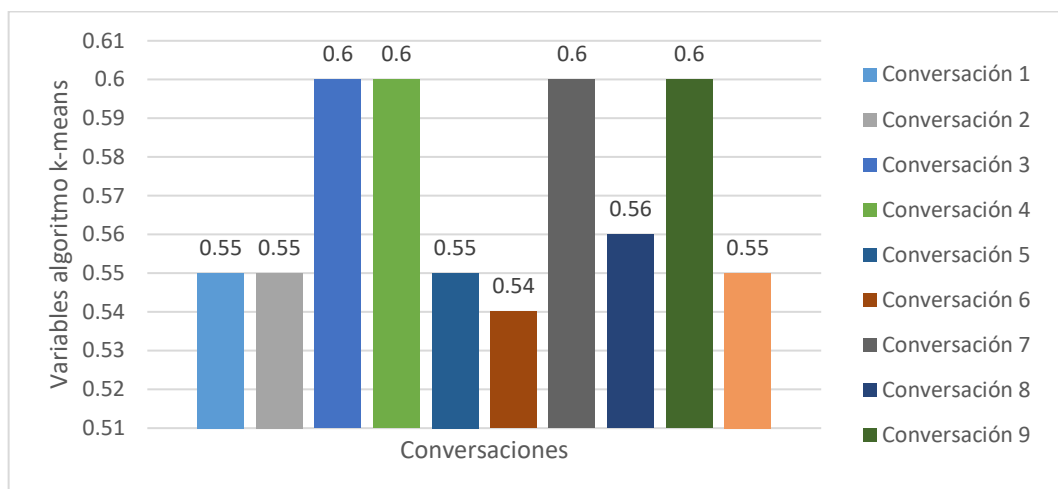


Figura 21. Distribución de resultados en personas normales

Fuente: El Autor

Elaboración: El Autor

5.1.1.1 Análisis de los resultados obtenidos por el algoritmo k-means en personas con daño cognitivo.

En este apartado analizaremos los resultados obtenidos para tres conversaciones con personas con daño cognitivo, mediante el método de agrupamiento cluster y el algoritmo k-means detallada en el apartado 4.1.3.1.2. Para cada uno de ellos, se muestra una tabla con los valores obtenidos en los parámetros de voz y la tasa de acierto.

5.1.1.1.1 Conversación 1.

A continuación, describimos el análisis que se realiza en la conversación 1 agrupando 2 cluster con el algoritmo k-means.

Tabla 8. Reconocimiento de emociones con configuración de 2 cluster

Variables	Cluster		Tasa de acierto
	1	2	
emoción	Tristeza	alegría	0.53
Pitch (Hz)	132.57	159.12	
Jitter %	3.1	2.4	
Shimmer %	17.78	19.06	
Intensidad (dB)	78.35	79.18	

Fuente: El Autor

Elaboración: El Autor

La figura 22, muestra cómo están distribuidos los valores de las variables pitch, jitter, shimmer e intensidad según la emoción de tristeza y alegría mostrados en la Tabla 8.

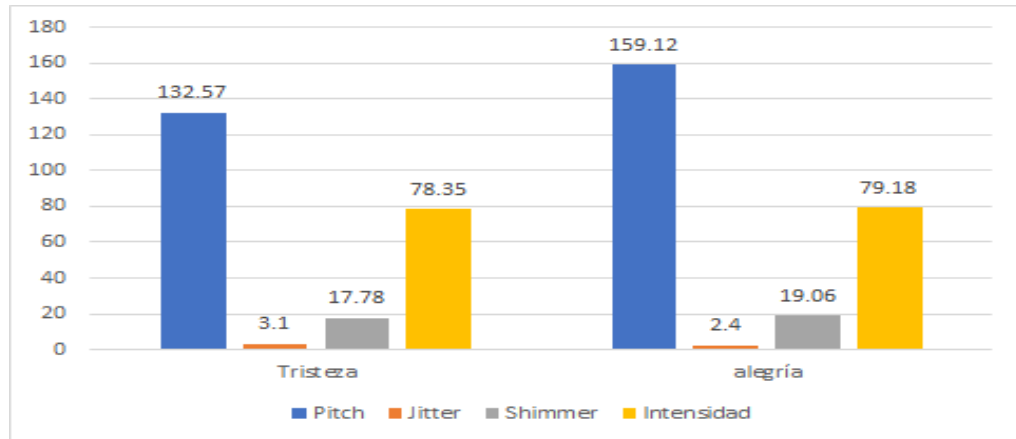


Figura 22. Distribución de las emociones de tristeza y alegría mostradas en la tabla 8
Fuente: El Autor
Elaboración: El Autor

En la conversación 1 se realiza un análisis con 2 cluster, lo cual dio una tasa de aciertos del 0.53% en el reconocimiento de emociones, siendo la alegría y tristeza las que mejor se identifican.

En esta tabla 8 observamos que:

- En esta conversación, la tasa de acierto de las emociones es mayor al realizar cluster a través de dos grupos tristeza y alegría con un porcentaje del 0.53%.
- A través de los diferentes parámetros de voz, la alegría siempre va a tomar valores de pitch e intensidad superiores a los de la tristeza.

Se realiza un segundo análisis en la conversación 1, se toma en cuenta 3 cluster con una tasa de acierto del 0.50 %, como se muestra en la tabla 9.

Tabla 9. Reconocimiento de emociones aplicando cluster 3 en la conversación 1

Variables	Cluster			Tasa de acierto
	1	2	3	
emoción	sorpresa	alegre	tristeza	0.5
Pitch (Hz)	139.33	162.58	127.6	
Jitter %	1.99	2.5	5.02	
Shimmer %	17.16	19.16	19.4	
Intensidad (dB)	79	78.66	79.4	

Fuente: El Autor
Elaboración: El Autor

La figura 23, muestra cómo están distribuidos los nuevos valores de las variables del pitch, jitter, shimmer e intensidad con una nueva emoción que es la sorpresa.

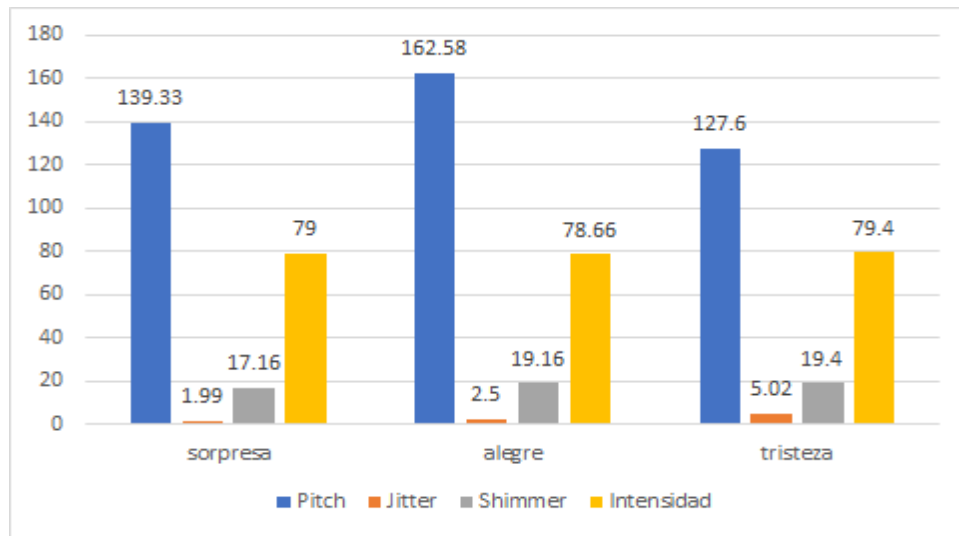


Figura 23. Distribución de la emoción sorpresa, alegre y tristeza
Fuente: El Autor
Elaboración: El Autor

En la Tabla 9 se realiza una agrupación de 3 cluster, se puede identificar una emoción más que es la sorpresa.

- La tasa de acierto realizando una clasificación con 3 grupos de emociones es aceptable con un porcentaje del 0.50%.
- En la emoción de la tristeza la perturbación de la intensidad que es el Shimmer es mayor a las demás emociones identificadas como son la sorpresa y alegría.
- De igual manera el parámetro Jitter que caracteriza la perturbación del pitch es mayor en la emoción de la tristeza con un porcentaje del 5.02% sobre la alegría 2.5% y sorpresa 1.99%.
- La emoción de la alegría tiene un elevado pitch sobre la tristeza y sorpresa.

5.1.1.1.2 Conversación 5.

En la conversación 5 se analizan los resultados obtenidos con el agrupamiento de 5 cluster, ya que existe un gran contenido emocional.

Tabla 10. Reconocimiento emociones aplicando 5 cluster en la conversación

Variable	CLUSTER					Tasa de acierto
	1	2	3	4	5	
emoción	alegre	triste	enfado	miedo	sorpresa	0.56
Pitch (Hz)	182.91	168.91	295.35	162.61	198.46	
Jitter %	1.8	2.3	3.69	2.84	1.62	
Shimmer %	16.04	19.75	17.77	15.45	12.57	
Intensidad (dB)	57.87	60.29	57.77	56.69	62.79	

Fuente: El Autor

Elaboración: El Autor

En la figura 24, muestra la distribución de las variables del pitch, jitter, shimmer, intensidad en las emociones de alegría, tristeza, enojo, miedo y sorpresa.

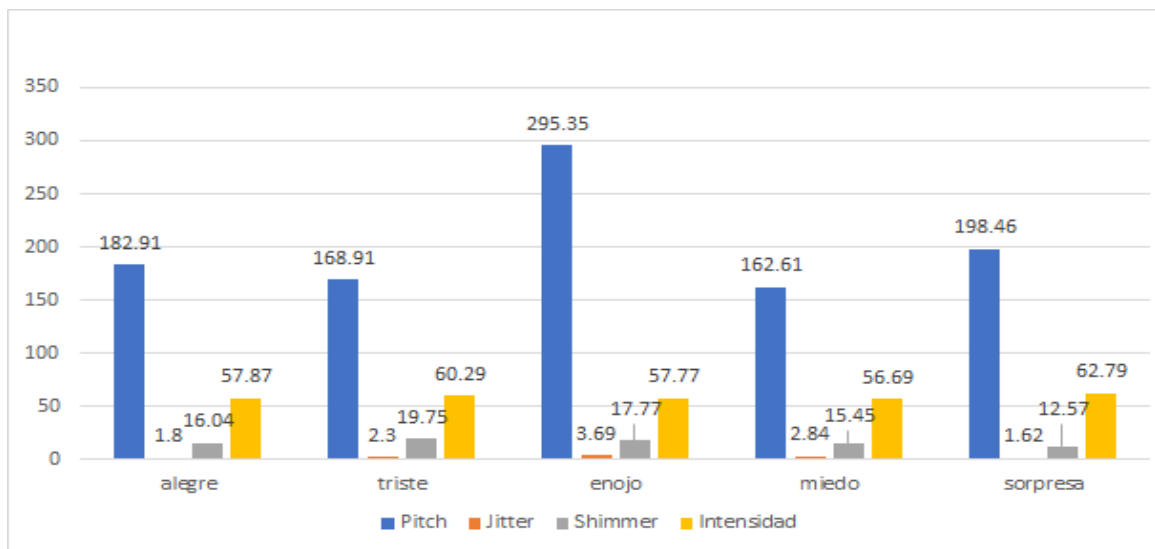


Figura 24. Distribución de las emociones alegría, tristeza, enojo, miedo y sorpresa

Fuente: El Autor

Elaboración: El Autor

En la Tabla 10 observamos que:

- Al igual que los análisis anteriores, existe una gran diferencia de rango entre la emoción de la alegría y tristeza, de igual manera de la sorpresa con la tristeza.
- La emoción del enfado es la que mayor aumento del pitch tiene sobre las demás emociones con un porcentaje del 295.35%.
- Este es el único caso, en los que la emoción de la sorpresa tiene valores más altos que la alegría.

- La emoción del miedo tiene un pitch (162.61 Hz) e intensidad (56.69 dB) inferior a las demás emociones.
- La sorpresa nunca se confunde con el enfado, ni con la tristeza.
- La alegría en este caso se puede observar que tiene a confundirse con la sorpresa.

5.1.1.1.3 Conversación 6.

En el análisis de la conversación 6 se realiza un análisis con 3 cluster.

Tabla 11. Reconocimiento de emociones aplicando 3 cluster

Variables	Cluster			Tasa de acierto
	1	2	3	
emoción	alegre	sorpresa	triste	
Pitch (Hz)	201.26	147.66	117.76	0.55
Jitter %	3.3	2.03	3.8	
Shimmer %	20.14	15.78	22.28	
Intensidad (dB)	51.63	55.8	51.34	

Fuente: El Autor

Elaboración: El Autor

La figura 25 muestra cómo se distribuyen los resultados obtenidos de las variables del pitch, jitter, shimmer e intensidad en las emociones de alegría, sorpresa y tristeza tal como se mencionan en la tabla 11.

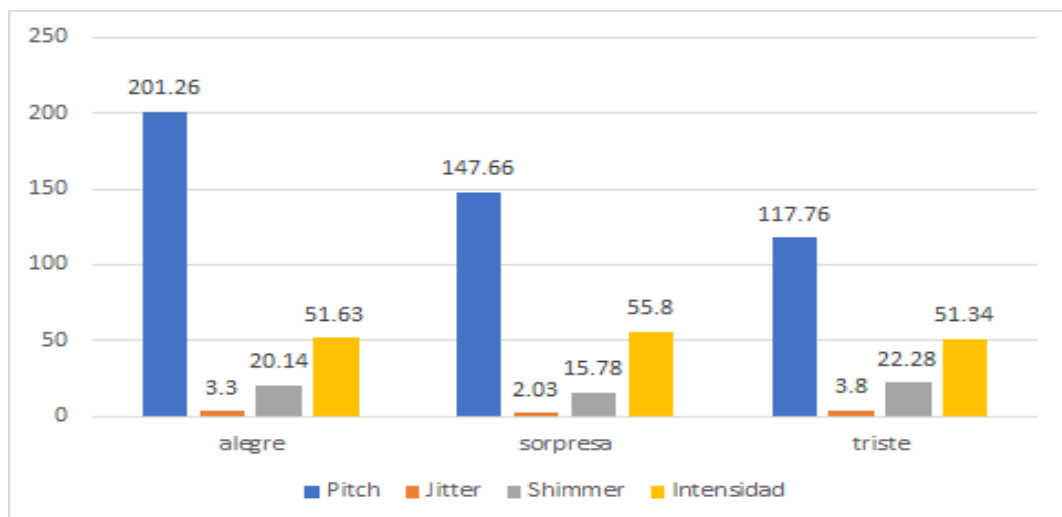


Figura 25. Distribución de las emociones de alegría, sorpresa y triste en la conversación 6

Fuente: El Autor

Elaboración: El Autor

En la conversación 6 se realiza un análisis con 3 cluster, lo cual dio una tasa de aciertos del 0.55% en el reconocimiento de emociones, siendo la alegría, tristeza y sorpresa las que mejor se identifican.

Observando la Tabla 11 llegamos a las siguientes conclusiones:

- La emoción que mejor se identifica es la tristeza, seguida de la alegría y la sorpresa.
- Hay ciertas emociones que nunca se confunden entre ellas
- La sorpresa nunca se confunde con la tristeza y la tristeza nunca se confunde con la sorpresa, ni tampoco con la alegría.
- En la emoción de Alegría el pitch siempre tiende a tomar valores elevados a diferencia de la tristeza que los valores del pitch son totalmente inferiores.

5.1.1.2 Análisis de los resultados obtenidos mediante el algoritmo k-means en personas sanas.

Analizamos los resultados obtenidos en tres conversaciones con los participantes sanos, con el método de agrupamiento cluster y el algoritmo k-means detallada en el apartado 4.1.3.1.2. Para cada uno de ellos, se muestra una tabla con los resultados obtenidos en los parámetros de voz y la tasa de acierto.

5.1.1.2.1 Conversación 2.

A continuación, se describe el análisis en la conversación 2 de un participante sano con el agrupamiento de 3 cluster con el algoritmo k-means.

Tabla 12. Reconocimiento de emociones aplicando 3 cluster en la conversación 2

Variables	Cluster			Tasa de acierto
	1	2	3	
emoción	neutro	tristeza	alegría	0.55
Pitch (Hz)	175.72	136.77	192.02	
Jitter	2.92	3.94	4.46	
Jitter %	12.86	19.62	18.97	
Shimmer %	74.79	72.26	72.78	

Fuente: El autor

Elaboración: El Autor

En la figura 26 podemos observar cómo se distribuyen los valores en las emociones de neutro, tristeza y alegría.

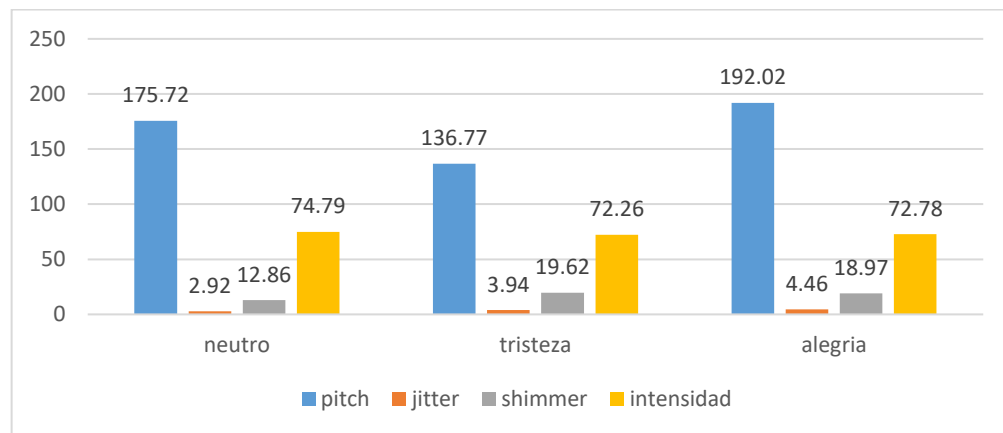


Figura 26. Distribución de las emociones de neutro, tristeza y alegría

Fuente: El Autor

Elaboración: El Autor

Observando la Tabla 12 y la figura 24 podemos analizar los siguiente:

- La emoción de la alegría es la que mayormente se reconoce en la conversación.
- El pitch en la emoción de la alegría es mayor a las emociones de neutro y tristeza.
- La tristeza nunca se confunde ni con la alegría y la neutra ya que su pitch e intensidad es siempre bajo.
- Con un agrupamiento de 3 grupos nos arroja una tasa de acierto del 0.55% que es aceptable en el reconocimiento emocional del habla.

Se realiza un segundo análisis, aplicamos el algoritmo k-means con un agrupamiento de 4 cluster.

Tabla 13. Reconocimiento de emociones aplicando 4 cluster en la conversación 2

Variables	Cluster				Tasa de acierto
	1	2	3	4	
Emoción	neutro	tristeza	alegre	sorpresa	0.55
Pitch (Hz)	175.72	136.77	195.83	190.39	
Jitter %	2.92	3.94	4.76	4.76	
Shimmer %	12.86	19.62	16.2	20.16	
Intensidad (dB)	74.79	72.26	77.08	70.94	

Fuente: El Autor

Elaboración: El Autor

En la figura 27 se observa la distribución de las 4 emociones identificadas a través de la configuración para 4 grupos del cluster.

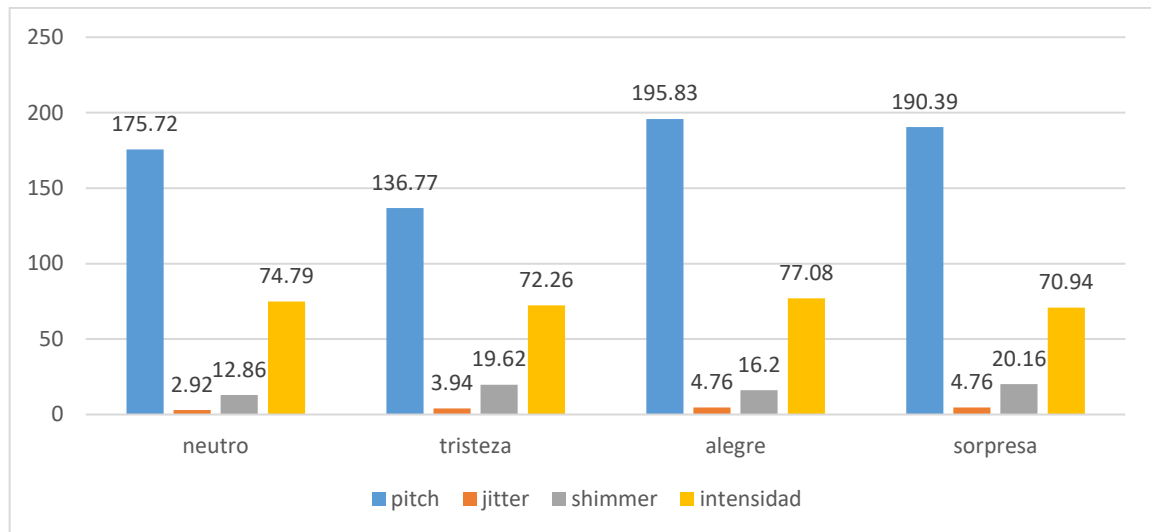


Figura 27. Distribución de las emociones de neutro, tristeza alegre y sorpresa
Fuente: El Autor
Elaboración: El Autor

Observando el resultado obtenido en la Tabla 13, podemos concluir lo siguiente:

- Se identificó la emoción de sorpresa, con una configuración de 4 cluster.
- Las emociones tanto como la sorpresa y alegría tienen un pitch alto con un porcentaje para la sorpresa de 190.39 Hz y alegría 195.83 Hz.
- La emoción de la alegría sigue mostrando mayor pitch e intensidad sobre las demás emociones.
- La tristeza es la que menor pitch tiene con un porcentaje de 136.77 Hz.
- La sorpresa tiene un pitch e intensidad mayor que la emoción neutra.

5.1.1.2.2 Conversación 7.

Se realiza el análisis de la conversación 7 con el algoritmo k-means con un numero de cluster igual a 3.

Tabla 14. Reconocimiento de emociones configurando 3 cluster en la conversación 7

Variables	Cluster			Tasa de acierto
	1	2	3	
emoción	neutro	sorpresa	Alegre	0.6
Pitch (Hz)	169.42	171.19	227.77	
Jitter %	5.1	2.99	4.24	
Shimmer %	23.23	20.85	18.79	
Intensidad (dB)	53.56	56.61	57.75	

Fuente: El Autor
Elaboración: El Autor

En la figura 28 se observa de mejor manera la distribución de las 3 emociones identificadas a partir del agrupamiento de 3 cluster con el algoritmo k-means en los diferentes parámetros de la voz.

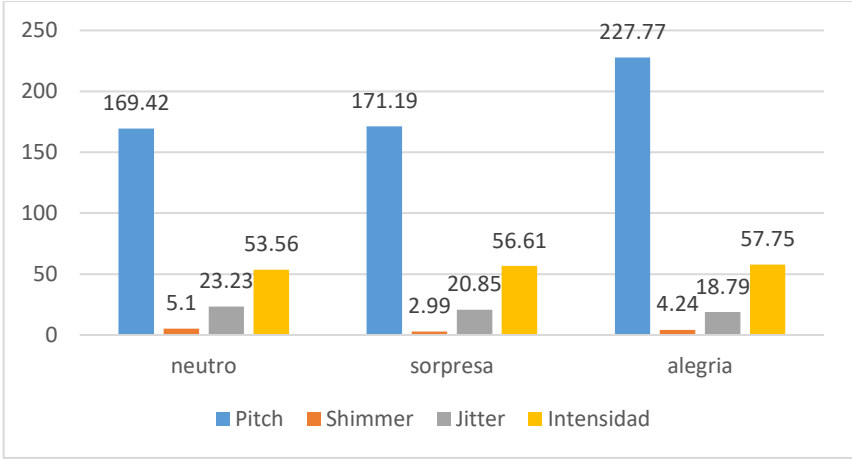


Figura 28. Distribución de las emociones neutro, sorpresa y alegría
Fuente: El Autor
Elaboración: El Autor

En Tabla 14 y la figura 28 observamos que:

- Se identifican tres grupos de emociones claras que son neutro, sorpresa y alegría.
- En la conversación 7 la emoción de la sorpresa tiene un pitch 171.19 Hz e intensidad 56.61 dB mayor a la emoción neutra 169.42 Hz y 53.56 dB.
- El Shimmer en la emoción neutra muestra una mejor estabilidad vocal del 5.1% sobre las emociones de sorpresa y alegría.

5.1.1.2.3 Conversación 10.

En la conversación 10 se realiza el análisis con el algoritmo k-means con una configuración del cluster que es igual a 3.

Tabla 15. Reconocimiento de emociones configurando 3 cluster en la conversación 10

Variables	Cluster			Tasa de acierto
	1	2	3	
emoción	neutro	alegre	enfado	0.55
Pitch (Hz)	228.79	233.89	269.15	
Jitter %	3.28	3.94	3	
Shimmer %	17.49	20.91	17.55	
Intensidad (dB)	70.42	68.94	72.16	

Fuente: El Autor
Elaboración: El Autor

En la figura 29 se observa la distribución de las 3 emociones neutro, alegre y enfado identificadas a partir del agrupamiento de 3 cluster con el algoritmo k-means.

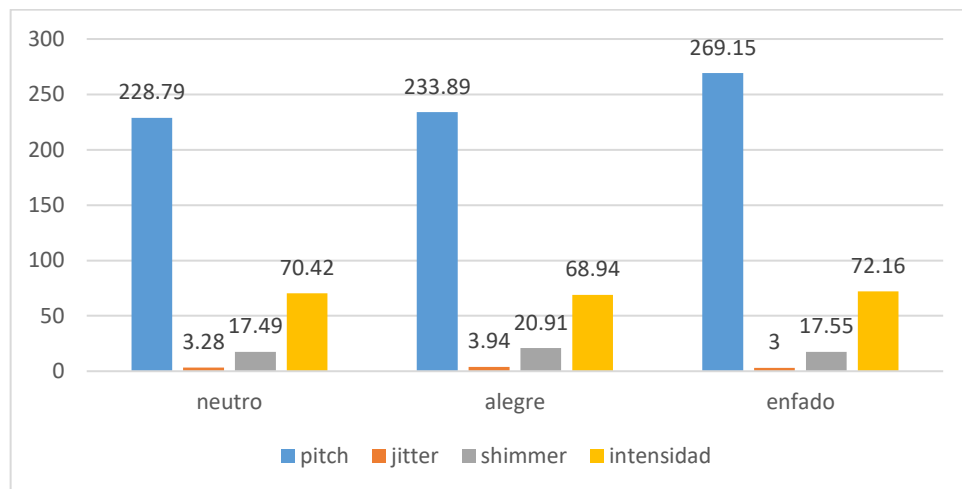


Figura 29. Distribución de las emociones neutro, alegre y enfado
Fuente: El Autor
Elaboración: El Autor

En la Tabla 15 y figura 29 observamos que:

- El enfado asocia valores de pitch (269.15 Hz) e intensidad (72.16 dB) muy altos con relación a la emoción de alegre y neutro.
- La emoción del enfado no se confunde con la neutra ya que existe un amplio rango de diferencia entre el pitch.
- La alegría y enfado pueden confundirse ya que toman valores similares en los diferentes parámetros de voz.
- En la conversación 10 se ha obtenido una tasa de aciertos del 0.55 % que es un valor aceptable en el reconocimiento emocional a través del habla.

En el (ver Anexo D) se muestra el análisis de las demás conversaciones en el reconocimiento emocional en el habla a través del algoritmo k-means.

5.1.1.3 Conclusiones del análisis.

Se han realizado distintos análisis a partir de la base de datos emocional con dos grupos distintos de personas. La tasa de acierto en personas sanas es del 0.57% y personas con daño cognitivo es del 0.49%. La tabla 16 muestra los resultados.

Tabla 16. Tasa de acierto en personas sanas y con daño cognitivo

Participantes	Tasa de acierto
Sanos	0.57
Daño Cognitivo	0.49

Fuente: El Autor

Elaboración: El Autor

- Se concluye que la tasa de aciertos en el reconocimiento de emociones en el habla a través del método de agrupación cluster con la selección del algoritmo k-means, en ambos grupos de participantes es aceptable.
- Los parámetros de voz como el pitch y la intensidad muestran valores relacionados a una emoción determinada.
- La emoción que no se identifica en los diferentes análisis es el asco y en pocas ocasiones el miedo.
- En las conversaciones realizadas con personas sanas se identifica en gran manera la emoción neutra por la manera espontánea que se lleva la conversación.
- En todos los análisis anteriores la emoción de alegría nunca se confunde con la tristeza y viceversa.
- Los niveles de reconocimiento alcanzados están afectados por factores como la calidad y claridad del material audiovisual, la duración de la conversación y la identificación de las emociones por parte del observador.

Tras completar el proceso de grabación, el resultado obtenido es una base de datos en donde quedan guardados en diferentes archivos los datos correspondientes a las grabaciones en audio y video. Para las grabaciones de audio, se tiene información sobre los resultados obtenidos del análisis de los parámetros de voz realizado por la herramienta Praat, así como la evaluación a través del método cluster. La información se encuentra disponible en el siguiente repositorio para futuras investigaciones en el reconocimiento emocional: <https://git.taw.utpl.edu.ec/drvaldivieso/Base-de-datos-emocional-multimodal.git>

CONCLUSIONES

Al finalizar con el presente trabajo de fin de titulación se concluye con lo siguiente:

- Las bases de datos emocionales resultan muy útiles a la hora de desarrollar sistemas de computación emocional. Son utilizadas principalmente en sistemas de reconocimiento de emociones, pero también pueden usarse para desarrollar sistemas de síntesis de emociones.
- De acuerdo con los resultados obtenidos al aplicar el algoritmo de agrupamiento k-means, se podría afirmar que la base de datos ha sido evaluada en el ámbito del reconocimiento del habla emocional y que la tasa de acierto son valores aceptables para ser utilizado en otros estudios de investigación relacionados con el análisis de parámetros de voz y la emoción.
- A través de los parámetros de voz como son el Pitch y la Intensidad se puede identificar emociones básicas como la alegría, tristeza, enfado y neutro, ya que existe una gran diferencia de rangos numéricos entre las emociones.
- Los patrones emocionales que más prevalecen en las personas entrevistadas que tienen daño cognitivo son “alegría, tristeza y enojo”, las mismas pueden ser validadas a través de un experto humano.
- Las emociones que mejor se agrupan y reconocen son la tristeza, el enfado y la neutra; y las que peor, la alegría y la sorpresa. Éstas últimas son emociones positivas que se confunden entre ellas.

RECOMENDACIONES

En base al presente trabajo de fin de titulación, se proponen las siguientes recomendaciones:

- Incrementar la cantidad de muestras de la base de datos emocional, realizando grabaciones en otros hogares del país a personas con algún tipo de daño cognitivo.
- Ampliar el número de parámetros de voz para realizar una mejor diferenciación de los grupos estudiados en este trabajo.
- Diseñar una herramienta automática para identificar deterioro cognitivo mediante procesamiento digital de voz.
- Plantear nuevos métodos que faciliten la tarea de clasificación y análisis de los datos en la base de datos creada.

BIBLIOGRAFÍA

- Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing*, 6(3), 209–222. <http://doi.org/10.1109/TAFFC.2015.2392932>
- Abreu, M., Deterioro, I., & Moderado, C. (2015). *Trabajo de diploma* “. Universidad Central “Marta Abreu” de Las Villas. Retrieved from [http://dspace.uclv.edu.cu/bitstream/handle/123456789/3117/Alex Romero Baluja.pdf?sequence=1&isAllowed=y](http://dspace.uclv.edu.cu/bitstream/handle/123456789/3117/Alex_Romero_Baluja.pdf?sequence=1&isAllowed=y)
- Alejandro, C., Valencia, T., Alexánder, M., & López, Á. (2013). Índice General.
- Álvarez, A., Cearreta, I., López, J. M., Arruti, A., Lazkano, E., Sierra, B., & Garay, N. (2007). in the Automatic Emotion Recognition Using Feature Subset Selection Based on Evolutionary Algorithms, 423–430.
- Álvarez, A., Sierra, B., Arruti, A., Lópezgil, J. M., & Garay-Vitoria, N. (2015). Classifier subset selection for the stacked generalization method applied to emotion recognition in speech. *Sensors (Switzerland)*, 16(1), 1–26. <http://doi.org/10.3390/s16010021>
- Bastidas Mendieta, P. V. (2017). La Universidad Católica de Loja.
- Bell, J. (2014). *Hands-On for Developers and Technical Professionals. Machine Learning: Hands-On for Developers and Technical Professionals.*
- Causa, E., & Sosa, A. (2007). La computación afectiva y el arte interactivo. *Área Transdepartamental de Artes Multimediales*, 52.
- Costantini, G., Iaderola, I., Paoloni, A., & Todisco, M. (2014). EMOVO Corpus: an Italian Emotional Speech Database. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 3501–3504. Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/pdf/591_Paper.pdf
- Ekman & Friesen, 1980 Facial-Sign-Of-Emotional-Experience.pdf. (n.d.).
- Espinosa, H. P., & García Reyes, C. A. (2010). Reconocimiento de Emociones a Partir de Voz Basado en un Modelo Emocional Continuo, (Ccc).
- General, C. de S. (2012). Diagnóstico y Tratamiento del Deterioro cognitivo en al Adulto Mayor en el Primer Nivel de Atención. *Catalogo Maestro De Guis De Practicas Clinicas*, 1, 8–18.
- Gomez Bermeo, P. F. (2017). Análisis y Comparación del índice kappa para la identificación de patrones en personas ancianas.
- Jose Francisco Solís Villareal. (2011). *Modelo de procesamiento de voz para la clasificación de*. Instituto Politécnico Nacional.

<http://www.repositoriodigital.ipn.mx/bitstream/handle/123456789/7767/tesis.pdf?sequence=1>

- José Manuel Molina López, J. G. H. (2004). Técnicas de análisis de datos, 160.
- Koelstra, S. (2012). Deap: A database for emotion analysis; using physiological signals. *Affective Computing*, ..., 3(1), 18–31. <http://doi.org/10.1109/T-AFFC.2011.15>
- Llorente, C. R., Manuel, D. J., & Martínez, M. (2017). Diseño, implementación y evaluación de técnicas de identificación de emociones a través de la voz.
- Lógica, E. De, De, F., Promesas, L., & Computación, D. (2015). Máster Universitario en Filosofía Teórica y Práctica Emociones : Calculemos.
- López, J. M., Cearreta, I., Garay, N., & ... (2006). Creación de una base de datos emocional bilingüe y multimodal. *Proceedings of the 7th*
- Manuel, C., & Sánchez, M. (n.d.). *Modelado de la cualidad de la voz para la síntesis del habla expresiva*. Universitat Ramon Llull.
- Margarita, M. (2016). La Universidad Católica de Loja.
- Ortego Resa, C. (2009). *DETECCION DE EMOCIONES EN VOZ ESPONTÁNEA*. Universidad Autonoma de Madrid.
- Ringeval, F., Sonderegger, A., Sauer, J., & Lalanne, D. (2013). Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, (i). <http://doi.org/10.1109/FG.2013.6553805>
- Santiago Planet Garcia, Jose Antonio Morán Moreno, L. F. F. (2006). *Reconocimiento de emociones basado en el análisis de la señal de voz parametrizada*. Retrieved from <http://users.salleurl.edu/~si05993/material/pdf/planet06.pdf>
- Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1), 42–55. <http://doi.org/10.1109/T-AFFC.2011.25>
- Tobergte, D. R., & Curtis, S. (2013). No Title No Title. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <http://doi.org/10.1017/CBO9781107415324.004>

ANEXOS

Anexo A: Estructura de conversación.



Universidad Técnica Particular de Loja

Estructura conversación personas adultos mayores con daño cognitivo.

Saludo.

¿Cómo se encuentra?

¿Con quién vive?

Si convive con alguien, preguntar por nombres de los familiares (hijos, hermanos, padres).

¿En qué estado civil se encuentra ahora?

Cuénteme sobre sus hábitos alimenticios

¿Qué trabajo realiza actualmente?

Actividades y gustos varios del participante.

Agradecimientos.

Anexo B: Petición ingreso al centro del adulto mayor de Loja



Universidad Técnica Particular de Loja

PATRONATO DE AMPARO
SOCIAL MUNICIPAL
AUTORIZADO



Tania García Córdova
PRESIDENTA

Jug. Córdova

Loja, 13-07-2016

Abg. Mag.

Tania García Córdova

DIRECTORA DEL PATRONATO DE AMPARO SOCIAL MUNICIPAL DE LOJA.

De mi consideración

Distinguida Dra., Deseándole éxitos en las funciones que acertadamente dirige. Yo, Daniel Valdivieso Alarcón, estudiante de la carrera de Sistemas Informáticos y Computación de la Universidad Técnica Particular de Loja quiero pedirle a usted la autorización para el ingreso al Centro del Adulto Mayor, ya que como uno de los requisitos de mi proyecto de fin de titulación necesito realizar una conversación con los adultos mayores en especial con aquellos que tengan un daño cognitivo, acerca de cómo ha sido su vida al transcurso del tiempo y su bienestar personal actual, la misma que debe quedar plasmada en filmaciones, recalco que el único fin del video es académico, y tiene como finalidad la obtención de mi Título de Tercer Nivel, puesto que el tema de mi tesis, es "Creación de una base de datos emocional multimodal de personas cognitivamente normales y con daño cognitivo". La conversación será exclusivamente entre los adultos mayores en especial con aquellos con un daño cognitivo, además de esto le doy a conocer que las preguntas serán solo acerca de la vida de los adultos las mismas que serán grabadas con una cámara digital.

Así mismo me comprometo a entregar una copia de mi proyecto de fin de titulación al Patronato de Amparo Social Municipal de Loja, para el uso pertinente del mismo.

Por la favorable atención a la misma, le anticipo mis más sinceros agradecimientos.

Fr.

Daniel Valdivieso Alarcón,
Estudiante UTPL

Fr.

Msc. Susana Arias Tapia
Docente Investigador UTPL

Pd. Adjunto Estructura Conversación y Consentimiento Informado

He. Guacilola Arriegas
Corred.

PATRONATO DE AMPARO
SOCIAL MUNICIPAL
PLANIFICACIÓN Y PROYECTOS

RECIBIDO: Luis Santa
FECHA: 13-06-2016
HORA: 16:45

Anexo C: Consentimiento Informado

CONSENTIMIENTO INFORMADO

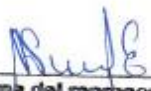
AUTORIZACIÓN INDIVIDUAL PARA LA INVESTIGACIÓN ACADÉMICA Y EL USO DE DATOS

Loja 23 de Julio del 2016

Yo, Gladys Esperanza Gracida Herrera, representante del
señoría los Angelica Herrera Poma autorizo al estudiante Daniel Rafael Valdivieso
Alarcón de la Carrera de Sistemas Informáticos y Computación de la UTPL para que use los
datos consignados en las filmaciones de entrevistas realizadas con el propósito de ejecutar el
trabajo de fin de titulación denominado "Creación de una base de datos emocional
multimodal de personas cognitivamente normales y con daño cognitivo"

Se me ha informado que la vida de mi representado/a no corre riesgo o peligro alguno, y que los
datos obtenidos son netamente con fines académicos, con la confidencialidad debida así como
también no será publicado ni difundido públicamente, en ningún medio de red social, de la misma
forma se me ha informado que mi representante puede retirarse de este estudio el momento que
considere adecuado.

Cualquier pregunta respecto a esta investigación la podre realizar al señor Daniel Rafael
Valdivieso Alarcón en forma personal o al celular 0990184993.


Firma del representante


Firma del Estudiante

Anexo D: Análisis de los resultados obtenidos por el algoritmo k-means en los participantes con daño cognitivo

En este anexo se muestra el análisis mediante cluster y la tasa de acierto en el reconocimiento emocional del habla.

Conversación 2.

Tabla 17. Reconocimiento de emociones configurando 3 cluster en la conversación 2

Variables	cluster			Tasa de acierto
	1	2	3	
Emoción	tristeza	alegría	miedo	0.26
Pitch	157.24	184.21	171.43	
Jitter	3.2	1.8	1.9	
Shimmer	16.10	13.08	17.33	
Intensidad	63.38	65.71	64.33	

Fuente: El Autor

Elaboración: El Autor

En la figura 30 se muestra la distribución de las emociones tristeza, alegría y miedo asociados a los valores numéricos del pitch, Jitter, Shimmer e intensidad.

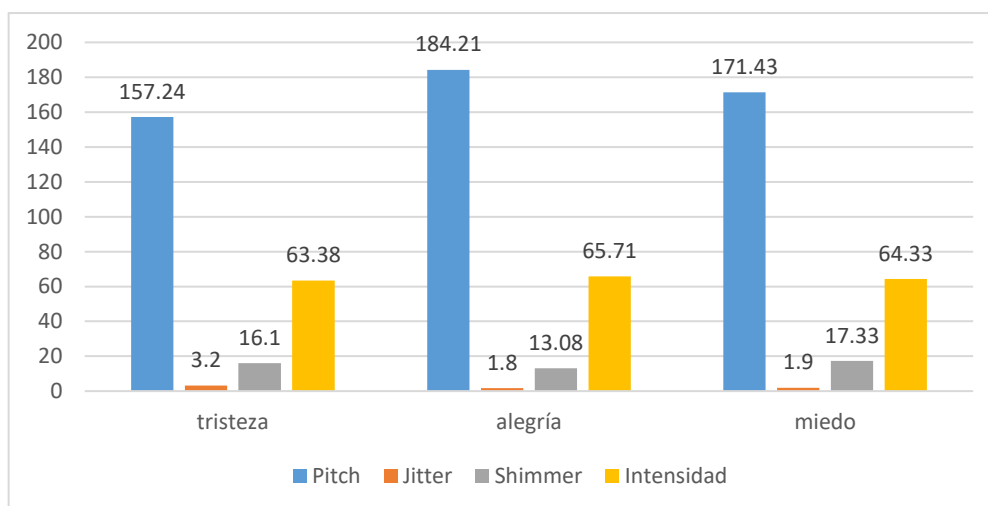


Figura 30. Distribución de las emociones tristeza, alegría y miedo

Fuente: El Autor

Elaboración: El Autor

En la tabla 17 y figura 30 podemos observar que:

- A través de la clasificación de 3 grupos de cluster y con los resultados obtenidos por el algoritmo k-means se puede identificar tres emociones que son la tristeza, alegría y miedo.

- La emoción de la alegría tiene un pitch de 184.21 Hz e intensidad de 65.71 dB que es mayor a las emociones de tristeza y miedo.
- La emoción del miedo tiene un pitch 171.43 e intensidad 64.33 dB mayor a la tristeza.

Conversación 3.

Tabla 18. Reconocimiento de emociones configurando 3 cluster en la conversación 3

Variables	cluster			Reconocimiento k-means
	1	2	3	
emoción	Tristeza	sorpresa	alegría	0.38
PITCH	147.3	151.1	183.45	
JITTER	1.93	2.44	1.78	
SHIMMER	11.53	14.08	10.79	
INTENSIDAD	66.44	68.13	70.98	

Fuente: El Autor

Elaboración: El Autor

La figura 31 muestra la distribución de la agrupación de las emociones de tristeza, sorpresa y alegría asociados a los valores numericos de los parámetros de voz.

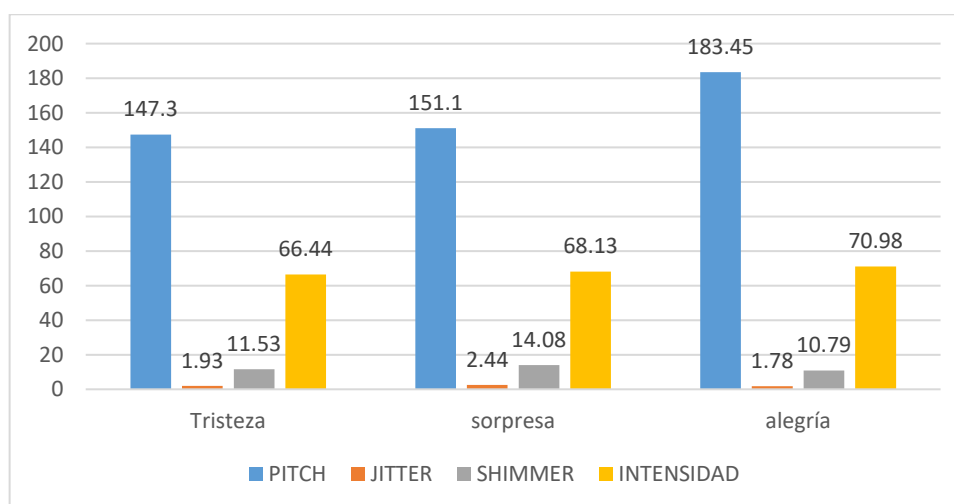


Figura 31. Distribución de las emociones tristeza, sorpresa y alegría

Fuente: El Autor

Elaboración: El Autor

En la tabla 18 y figura 31 podemos observar que:

Se realiza el análisis con un número de cluster igual a 3, no existe un agrupamiento correcto por lo cual como resultado se obtiene una tasa de reconocimiento emocional en el habla del 0.38% que es muy baja.

Conversación 4

Tabla 19. Reconocimiento de emociones configurando 5 cluster en la conversación 4

Variables	CLUSTER					Reconocimiento k-means
	1	2	3	4	5	
emoción	triste	Enfado	neutro	alegre	sorpresa	0.4
Pitch	115.48	120.72	116.45	121.43	116.79	
Jitter	1.73	1.58	2.3	1.55	4.31	
Shimmer	15.39	20.55	20.35	12.17	20.3	
Intensidad	53.85	54.62	58.32	55.99	53.31	

Fuente: El Autor

Elaboración: El Autor

La figura 32 muestra la distribución de 4 emociones básicas y su relación con los valores numéricos del pitch, Jitter, Shimmer e intensidad.

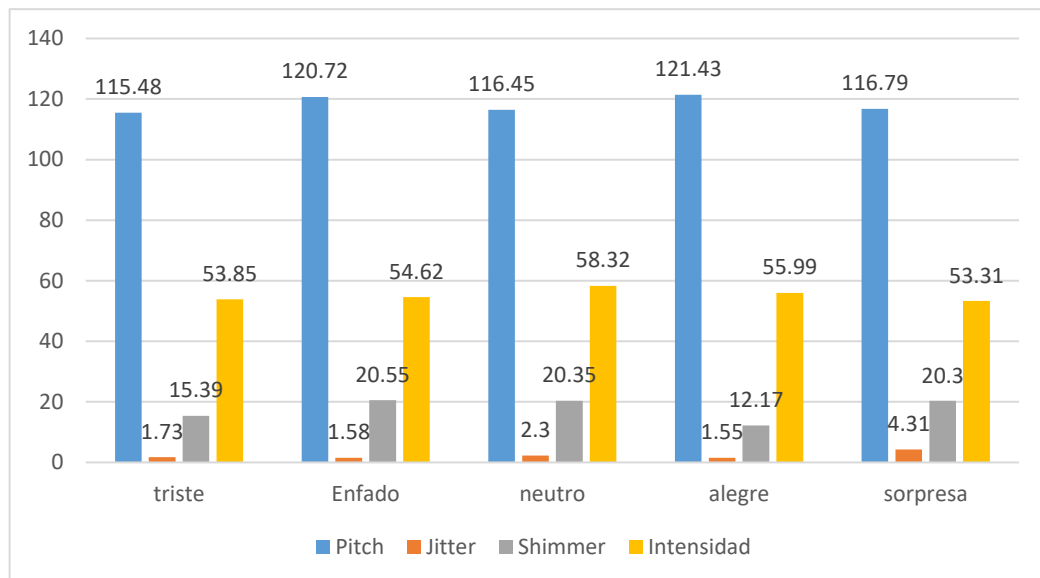


Figura 32. Distribución de las emociones tristeza, enfado, neutro, alegre y sorpresa

Fuente: El Autor

Elaboración: El Autor

En la tabla 19 y la figura 32 podemos observar que:

- Se realiza la clasificación con 5, el agrupamiento muestra indicadores a través de los parámetros de voz para identificar cada emoción, pero no existe un agrupamiento correcto por lo cual la tasa de acierto emocional es baja con un porcentaje del 0.40%.
- La emoción del enfado se puede confundir con la alegría por los valores similares que tienen en el pitch, Shimmer e intensidad.
- La sorpresa se confunde con la neutra.

- La tristeza se mantiene con valores de pitch e intensidad menor al enfado y alegría.

Conversación 7.

Tabla 20.Reconocimiento de emociones configurando 3 cluster en la conversación 7

Variables	CLUSTER			Reconocimiento
	1	2	3	
Emoción	alegre	triste	enfado	0.52
PITCH	201.81	120.72	266.69	
JITTER	3.56	2.59	3.93	
SHIMMER	20.49	19.92	17.28	
INTENSIDAD	70.13	67.93	71.08	

Fuente: El Autor

Elaboración: El Autor

La figura 33 muestra la distribución de las emociones alegre, triste y enfado cada una con los valores asociados del pitch, Jitter, Shimmer e intensidad.

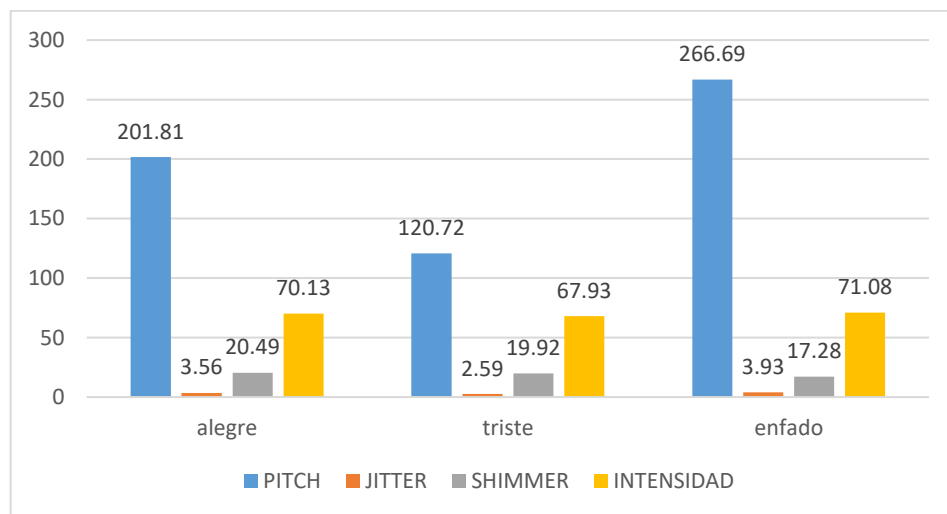


Figura 33.Distribución de las emociones alegre, triste y enfado

Fuente: El Autor

Elaboración: El Autor

En la tabla 20 y figura 33 podemos observar que:

- El enfado tiene un pitch de 266.69 Hz e intensidad dB mayor a las emociones de alegre y triste.
- Los valores bajos del pitch, Jitter e intensidad siempre se van asociar a una emoción de tristeza.
- El enfado nunca se confunde con la tristeza.

Conversación 8.

Tabla 21. Reconocimiento de emociones configurando 4 cluster en la conversación 8

Variables	CLUSTER				Reconocimiento k-means
	1	2	3	4	
emoción	enfado	alegría	sorpresa	triste	0.56
Pitch	209.34	181.05	153.23	97.68	
Jitter	2.72	4.74	2.78	5.8	
Shimmer	16.19	21.22	19.76	29.22	
Intensidad	80	78.54	76.36	74.72	

Fuente: El Autor

Elaboración: El Autor

La figura 34 muestra cómo se encuentra distribuido las emociones de enfado, alegría, sorpresa y triste con los valores obtenidos de los parámetros de voz.

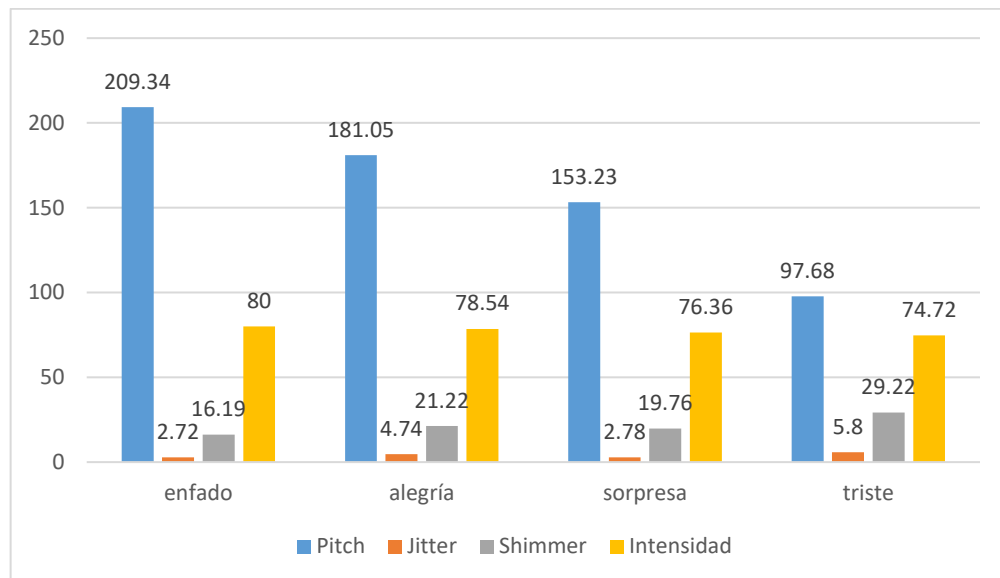


Figura 34. Distribución de las emociones enfado, alegría, sorpresa y triste

Fuente: El Autor

Elaboración: El Autor

En la tabla 21 y figura 34 podemos observar que:

- Al realizar un agrupamiento por 4 cluster se obtiene una tasa de acierto en el habla emocional del 0.56% que es aceptable.
- La alegría al igual que otros análisis sigue siendo la emoción que más valores mínimos asocia del pitch 97.69 e intensidad 74.72 en esta conversación.
- El enfado con un pitch 209.34Hz e intensidad de 80 dB, es la emoción que asocia más valores altos sobre las demás emociones.

Conversación 9.

Tabla 22. Reconocimiento de emociones configurando 3 cluster en la conversación 9

Variables	Cluster			Reconocimiento k-means
	1	2	3	
EMOCION	NEUTRO	ALEGRE	TRISTE	0.55
PITCH	133.78	148.98	123.52	
JITTER	1.95	2.53	2	
SHIMMER	17.45	22.57	19.99	
INTENSIDAD	73.82	71.15	69.92	

Fuente: El Autor

Elaboración: El Autor

La figura 35 muestra la distribución de las emociones neutro, alegre y triste con los valores asociados a los parámetros de voz.

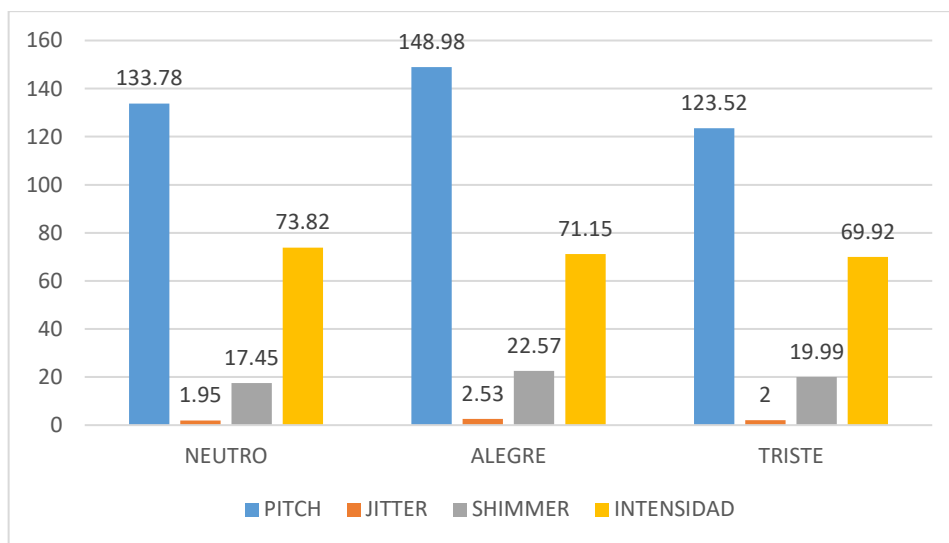


Figura 35. Distribución de las emociones neutro, alegre y triste

Fuente: El Autor

Elaboración: El Autor

En la tabla 9 y figura 35 podemos observar que:

- Se ha realizado un análisis de 3 cluster, se logra obtener una tasa de acierto del 0.55 % que es buena.
- La emoción de alegría nunca se confunde con la tristeza.
- La alegría tiene un pitch 148.98 mayor sobre las demás emociones, es un buen indicador al momento de reconocer las emociones.

Conversación 10.

Tabla 23. Reconocimiento de emociones configurando 3 cluster en la conversación 10

Variables	Cluster			Reconocimiento
	1	2	3	
EMOCION	alegre	sorpresa	tristeza	0.6
PITCH	163.83	156.86	131.65	
JITTER	3.53	2.68	3.19	
SHIMMER	19.69	15.75	17.67	
INTENSIDA	71.87	71.55	68.46	

Fuente: El Autor

Elaboración: El Autor

La figura 36 muestra la distribución de las emociones alegre, sorpresa y tristeza con los valores asociados a los parámetros de voz.

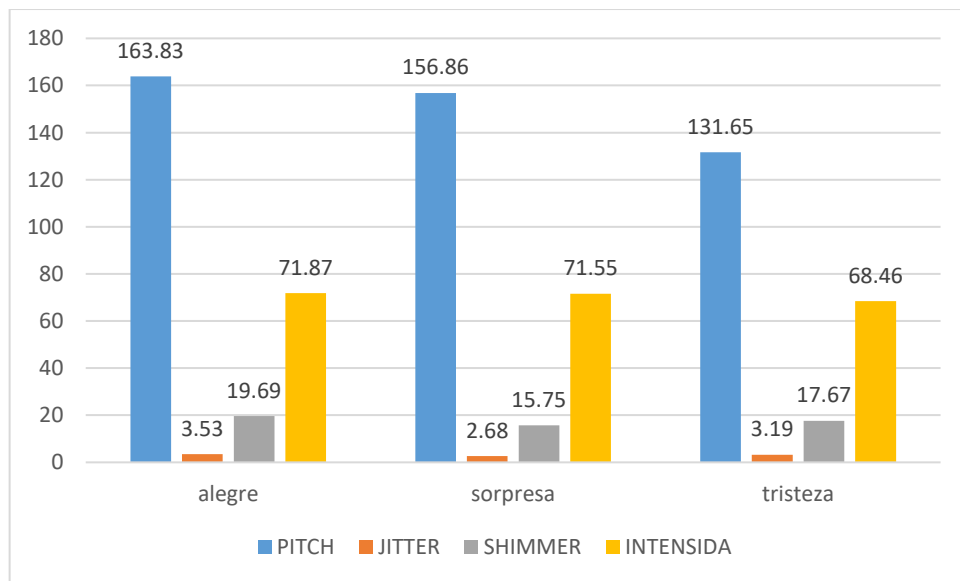


Figura 36. Distribución de las emociones alegre, sorpresa y tristeza

Fuente: El Autor

Elaboración: El Autor

En la tabla 10 y figura 36 podemos observar que:

- Se ha realizado un análisis de 3 cluster, se logra obtener una tasa de acierto del 0.60% que es buena.
- La emoción de la sorpresa y alegría pueden confundirse por los valores similares asociados en los diferentes parámetros de voz.
- La emoción de la alegría y tristeza nunca se confunden ya que tienen un rango diferenciado en los valores obtenidos de los parámetros de voz.

- La emoción de la sorpresa con un pitch 156.86 Hz e intensidad 71.55 es mayor a la emoción de tristeza pitch 131.65 Hz e intensidad 68.46 dB por lo cual nunca se van a confundir.