

UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE INGENIERÍA DE SISTEMAS

*Trabajo de titulación previo a
la obtención del título de
Ingeniera de Sistemas*

PROYECTO TÉCNICO:

**CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO
EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE
TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD
AUDITIVA.**

Autora:

Inés Patricia Yambay Aulla

Tutora:

Ing. Paola Cristina Ingavélez Guerra

Cuenca – Ecuador

2018

CESIÓN DE DERECHOS DE AUTOR.

Yo Inés Patricia Yambay Aulla con documento de identificación No. 010591753-8, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana, la titularidad sobre los derechos patrimoniales en virtud de que soy autora del trabajo de titulación: **CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD AUDITIVA**, mismo que ha sido desarrollado para optar por el título de *Ingeniera de Sistemas*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en condición de autora, reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, Diciembre del 2018



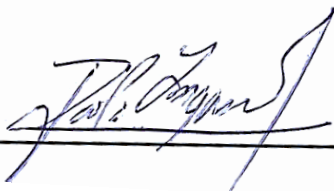
Inés Patricia Yambay Aulla

C.I.: 0105917538

CERTIFICACIÓN.

Yo, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD AUDITIVA**, realizado por Inés Patricia Yambay Aulla, obteniendo el *Proyecto Técnico*, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, Diciembre del 2018.



Ing. Paola Cristina Ingavéz Guerra

Tutora del trabajo de titulación

DECLARATORIA DE RESPONSABILIDAD.

Yo Inés Patricia Yambay Aulla con número de cédula 010591753-8, autora del trabajo de titulación: **CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD AUDITIVA**; certifico que el total contenido de este *Proyecto Técnico* es de mi exclusiva responsabilidad y autoría.

Cuenca, Diciembre del 2018.



Inés Patricia Yambay Aulla

C.I.: 0105917538

AGRADECIMIENTOS.

En primer lugar, quiero expresar mis más sinceros agradecimientos a la Ing. Paola Ingavélez quien ha sido mi tutora durante el desarrollo de este proyecto tecnológico, me ha brindado su tiempo, sus conocimientos y me ha guiado durante esta etapa.

Agradezco a la Universidad Politécnica Salesiana por abrirme sus puertas y brindarme una educación de calidad, también quiero agradecer todos los docentes que intervinieron durante mi proceso formativo y me brindaron sus conocimientos durante sus clases impartidas.

Finalmente, quiero agradecer a todos los que conforman la Catedra UNESCO “Tecnologías de Apoyo para la Inclusión Educativa” y su Grupo de Investigación en Inteligencia Artificial y Tecnología de Asistencia GI-IATa, al centro gerontológico “Rosa Elvira de León” por abrirme sus puertas y brindarme su ayuda para el desarrollo de este proyecto.

Inés Patricia Yambay Aulla

DEDICATORIA.

Al haber alcanzado la etapa más importante de mis estudios académicos quiero dedicar el desarrollo del presente trabajo primeramente a Dios quien me ha bendecido con salud, vida, sabiduría y una familia maravillosa, me ha dado paz en los momentos tormenta y ha guiado mi camino.

A mis padres quienes me han entregado su amor y apoyo de manera incondicional, agradezco su esfuerzo, sacrificio y velar por mi bienestar, por estar siempre a mi lado aconsejándome y enseñándome a no rendirme ante cualquier obstáculo, para mí son el tesoro más valioso que Dios me pudo dar y la vida no me alcanzaría para agradecerles por todo su amor y sacrificios. También quiero agradecer a mis hermanos y familia quienes me han aconsejado y animado a seguir adelante y cumplir con mis metas y objetivos.

Finalmente quiero dedicar este proyecto a mis amigos y amigas que he conocido a lo largo de mi vida, quienes me han escuchado y aconsejado, gracias por su apoyo y su amistad.

Inés Patricia Yambay Aulla

Índice de Contenido

RESUMEN.....	1
ABSTRACT.....	2
1. INTRODUCCIÓN.....	1
2. OBJETIVOS.....	3
2.1. General	3
2.2. Específicos.....	3
3. ESTADO DEL ARTE.....	4
3.1. La discapacidad Auditiva a nivel Mundial	4
3.1.1. Problemas asociados a la discapacidad auditiva.....	4
3.1.2. La discapacidad auditiva en el Ecuador	4
3.2. El Adulto Mayor	6
3.2.1. Adulto Mayor y Discapacidad	6
3.2.2. Adulto Mayor en Ecuador	7
3.3. Accesibilidad	7
3.3.1. Accesibilidad para personas con discapacidad auditiva post-lingüística - adultos mayores 8	
3.3.2. Subtitulado como apoyo para las personas con discapacidad auditiva.....	8
3.3.2.1. Pautas de accesibilidad para subtitulado.....	8
3.4. Fundamentos básicos de los Sistemas de reconocimiento de voz	9
3.4.1. Arquitectura de un sistema de reconocimiento de habla.....	9
3.4.1.1. Pre procesado del audio.....	9
3.4.1.2. Reconocimiento de Patrones.....	10
3.4.1.3. Decisión	11
3.4.2. Métodos que emplean.....	11
3.4.3. Modelos que emplea un sistema de reconocimiento del habla	13
3.4.4. Herramientas existentes.....	15
3.5. Trabajos Relacionados.....	16
4. ARQUITECTURA GENERAL DE LA PROPUESTA PLANTEADA.....	18
4.1. Diseño y Desarrollo de la aplicación de escritorio.....	18
4.1.1. Requerimientos para el Desarrollo del Sistema.....	18
4.1.2. Diseño y construcción de la arquitectura del sistema.	18
4.1.3. Especificaciones técnicas la aplicación desarrollada.....	18
4.1.3.1. Especificaciones de hardware	18
4.1.3.2. Especificaciones de software	19

4.1.3.3. Herramientas de software elegidas para el desarrollo	19
4.1.3.4. Python.....	19
4.1.3.5. Numpy.....	19
4.1.3.6. Matplotlib.....	20
4.1.3.7. CMUSphinx	20
4.1.3.8. Qt designer.....	20
4.1.3.9. PYQT4	20
4.1.3.10. Paquetes y librerías adicionales	21
4.1.3.11. Audio – pydub, pyaudio, sounddevice	21
4.1.3.12. Struct.....	21
4.1.3.13. Wave.....	21
4.1.3.14. Urllib3.....	21
4.2. Diagramas de Secuencia	21
4.3. Arquitectura general de la aplicación.....	22
4.3.1. Capa de Interacción	23
4.3.2. Capa de Sistema Inteligente.....	25
4.3.3. Base de Conocimiento	26
4.3.3.1. Entrenamiento y actualización del modelo Acústico y modelo de lenguaje.....	27
5. EXPERIMENTACIÓN Y RESULTADOS	30
5.1. Pruebas.....	30
5.2. Resultados obtenidos.....	33
5.2.1. Resultados de la evaluación de la aplicación en cada área.....	33
5.2.2. Resultados de la percepción del adulto mayor hacia la aplicación desarrollada	35
5.2.3. Resultados del nivel de comprensión del adulto mayor después de implementar la aplicación en su terapia de estimulación cognitiva	36
6. CONCLUSIONES Y TRABAJO FUTURO	36
6.1. Conclusiones.....	36
6.2. Recomendaciones.....	38
6.3. Trabajo futuro	39
REFERENCIAS	40
ANEXOS.....	43
• Modelo de encuesta aplicada	43
• Certificado de Participación en el IX Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2018) - VI Conferencia Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones para mejorar la Accesibilidad (ATICAcces 2018).....	47

- Consentimiento informado, firmado por la Psicóloga del Centro gerontológico “Rosa Elvira de León” 48

Índice de Figuras

Figura 1. Tipos de Discapacidad en Ecuador.....	5
Figura 2. Discapacidad en Ecuador, grados de discapacidad auditiva	5
Figura 3. Discapacidad en Ecuador, grupos etarios	6
Figura 4. Proyección de Tasas de Fecundidad 1970-2050 [1].....	7
Figura 5. Arquitectura de los modelos ocultos de Márkov.	12
Figura 6: Modelo de una neurona artificial o Perceptrón simple [19].	13
Figura 7. Procesamiento de señales para estructuras el modelo acústico [18].	14
Figura 8. Estructura del diccionario fonético español Ecuador.	14
Figura 9. Estructura del modelo de lenguaje.	15
Figura 10. Diagrama de Secuencia: Reconocimiento y transcripción del habla.....	22
Figura 11. Diagrama de Secuencia: Ejecución del entrenamiento	22
Figura 12. Arquitectura del Sistema	23
Figura 13. Pantalla Inicial de la Aplicación	24
Figura 14. Interfaz para el reconocimiento audio y transcripción a texto.	25
Figura 15. Interfaz para el entrenamiento del sistema.....	25
Figura 16. Configuración de los modelos utilizados por PocketSphinx.....	26
Figura 17. Funcionamiento e interacción de la base de conocimiento.....	26
Figura 18. Comando para extraer las características de los audios.	28
Figura 19. Comando para la adaptación y actualización de los modelos.	28
Figura 20. Pruebas del sistema de reconocimiento antes del entrenamiento del modelo acústico.....	31
Figura 21. Pruebas del sistema de reconocimiento después del entrenamiento del modelo acústico.....	31
Figura 22. Nivel de acierto de la aplicación en las áreas de estimulación cognitiva trabajadas.	34
Figura 23. Nivel de acierto de la aplicación en el área de orientación	34
Figura 24. Nivel de acierto de la aplicación en el área de lenguaje.	35
Figura 25. Nivel de acierto de la aplicación en el área de memoria y atención.....	35
Figura 26. Nivel de acierto de la aplicación en el área de funciones ejecutivas.....	35
Figura 27. Estadísticas sobre utilidad de la aplicación.	36
Figura 28. Estadísticas de implementación de la aplicación en las terapias.	36

Índice de Tablas

Tabla 1. Listado de frases con las que se trabajó en cada área.....	32
Tabla 2. Datos de las personas encuestadas.....	33

RESUMEN

En el transcurso de los últimos años hemos sido testigos del desarrollo acelerado de las nuevas tecnologías de información y comunicación (TIC's), hoy en día podemos destacar que han sido implementadas en todas las áreas de nuestra vida, con la finalidad de facilitar el desarrollo de actividades, simplificando procesos, automatizando tareas y facilitando el acceso a la información. Al hablar de tics debemos tener en cuenta que existen factores que pueden limitar su uso, las discapacidades sensoriales como visual y la auditiva se hace presente desde el nacimiento o se hace presente a medida que vamos envejeciendo y es una de las grandes limitantes a la hora de acceder a la información. Es por ello que es necesario garantizar que los recursos tecnológicos que se ofrecen estén diseñados de tal manera que sean accesibles para todas las personas ya sea que presenten o no algún tipo de discapacidad. Los adultos mayores se han visto beneficiados por este avance tecnológico, pues se han desarrollado y se siguen desarrollando sistemas de asistencia y apoyo que permiten mejorar su calidad de vida.

En el presente proyecto se desarrolla la propuesta planteada que tiene como finalidad apoyar el área de discapacidad y al adulto mayor, mediante el desarrollo de una aplicación de escritorio que facilita el acceso a la información.

El sistema desarrollado está conformado por una aplicación de escritorio que se encarga de capturar el habla emitida por el usuario, procesarla y extraer el subtítulo correspondiente, para ello hace uso de un sistema de reconocimiento de voz que analiza las frecuencias de audio recibidas, las compara con el modelo acústico entrenado previamente, el diccionario fonético y el modelo de lenguaje de todo el idioma español (Ecuador) y al finalizar el texto obtenido lo presenta en pantalla. Para entrenar al sistema se hizo una recopilación de frases, instrucciones y palabras que con más frecuencia se usan en un Centro Gerontológico durante el proceso de estimulación cognitiva, con la finalidad de que el subtítulo sea lo más claro y entendible posible. La interfaz del sistema permite que el usuario modifique tanto el contraste entre fondo y fuente y el tamaño de los caracteres con la finalidad de sea accesible para su público receptor.

ABSTRACT

Over the last few years we have witnessed the accelerated development of new information and communication technologies (ICT), nowadays we can distinguish that they have been implemented in all areas of our life, with the purpose of facilitating the development of activities, simplifying processes, automating tasks and facilitating access to information. When talking about tics we must take into account that there are factors that can limit their use, sensory disabilities as visual and auditory is present from birth or becomes present as we get older and is one of the major limitations at the time to access information. This is why it is necessary to ensure that the technological resources offered are designed in such a way that they are accessible to all people whether or not they have a disability. Older adults have benefited from this technological advance, since assistance and support systems have been developed and are being developed to improve their quality of life. In this project we develop the proposed proposal that aims to support the area of disability and the elderly

In the present project we developed the proposed approach that aims to support the area of disability and the elderly. through the development of a desktop application that facilitates access to information.

The developed system is made up of a desktop application that is responsible for capturing the speech issued by the user, processing it and extracting the corresponding subtitle, for this it makes use of a voice recognition system that analyzes the frequencies of audio received, compares them with the previously trained acoustic model, the phonetic dictionary and the language model all of the Spanish language (Ecuador) and at the end of the text obtained, he presents it on the screen. In order to train the system, a compilation of phrases, instructions and words was made that more often used in a Gerontological Center during the process of cognitive stimulation, with the purpose of subtitling being as clear and understandable as possible. The interface of the system allows the user to modify both the contrast between background and font and the size of the characters in order to be accessible to their audience.

1. INTRODUCCIÓN

A medida que vamos envejeciendo, pasamos por un deterioro físico y/o cognitivo, la percepción de nuestros sentidos se vuelve cada vez menos agudos, estos cambios sensoriales pueden provocar que la calidad de vida se vea afectada pues las actividades que antes se realizaban de forma rutinaria, ahora requieren de mayor refuerzo.

Según los datos recogidos en último censo realizado por el Instituto Ecuatoriano de Estadísticas y Censos (INEC) en el año 2010, en Ecuador existe una población de 1.2 millones de adultos mayores, de los cuales el 53.4% son mujeres, y el 46.6 % restante son hombres [1]. Siendo la esperanza de vida del hombre de 72 años y de la mujer de 78. Según las proyecciones realizadas para el año 2030 no habrá nacido suficiente gente que remplace su población, y según esto habrá más gente la que envejezca de la que nace.

La discapacidad auditiva, es la ausencia total o parcial de la percepción de audición en los oídos, esto puede darse desde el nacimiento en caso de ser congénitas, también pueden ser hereditarias o se puede dar en alguna etapa de la vida ya sea por una enfermedad, un accidente o el simple hecho de que estamos envejeciendo. La vejez es un territorio diverso que puede afectar de un modo distinto a cada individuo, pues en muchas de las ocasiones estas están acompañadas por el deterioro físico y/o cognitivo. Entre los problemas más frecuentes a los que se enfrenta el adulto mayor es la pérdida de audición la cual no solo afecta la capacidad de apreciar y entender el habla, sino que limita la intervención y contribución social, este último trae consigo depresión, disminución de la autoestima y aislamiento social. [2]

Tanto la pérdida de audición y la pérdida visual relacionada con el envejecimiento es irreversible pues está causada por la degeneración de las células sensoriales propias del envejecimiento. En la actualidad se dispone de auxiliares auditivos o audio prótesis y lentes, que pueden proporcionar la capacidad de oír y ver nuevamente en cualquier ambiente [3]. Entre las causas más frecuentes que pueden provocar un déficit auditivo en los adultos es la presbiacusia cuya prevalencia aumenta con la edad.

Con el auge de las nuevas tecnologías, se abre un nuevo camino lleno de posibilidades para mejorar el acceso a la comunicación, información, y más, sin olvidar que las herramientas tecnológicas deben eliminar las barreras que limiten la igualdad de acceso a la información; es por ello que, con la finalidad de facilitar su integración a través del uso de las nuevas tecnologías, estamos comprometidos a desarrollar herramientas que les permita crecer, aprender y ampliar horizontes.

Hoy en día existen diferentes apoyos tecnológicos, mismas que están evolucionando constantemente buscando mejorar la realidad que viven este grupo social, y ayudar a que puedan integrarse a una comunidad y/o sociedad. En el área de accesibilidad en contenidos audiovisuales para personas con discapacidad auditiva resaltamos que la

presencia del subtítulo debe ser indispensable pues el acceso a la información es un derecho y un requisito fundamental para la igualdad de oportunidades e inclusión [4].

Existen actualmente herramientas que permiten convertir un archivo de audio en texto, herramientas que generan subtítulo de manera automática a partir de videos (YOUTUBE), herramientas de reconocimiento de voz, etc. Sin embargo, estas herramientas en muchas de las ocasiones por no decir en todas no son gratuitas y dependen de una conexión a internet.

Es por esta necesidad que se plantea la idea de desarrollar el presente proyecto, con la finalidad de entregar una herramienta que genere subtítulo automático a partir del habla que se está dando en tiempo real por parte del locutor como un soporte para el adulto mayor que presenta algún tipo de pérdida de audición, y pueda acceder a la información y romper las barreras que impiden la accesibilidad.

Para el desarrollo de este sistema fue necesario la colaboración de la psicóloga de centro gerontológico, la cual nos brindó la información necesaria para armar nuestro corpus de datos con las frases más usadas durante las clases de estimulación cognitiva, y las instrucciones y/o indicaciones más utilizadas en el lugar por parte de las enfermeras. A partir de esta información se realizó sus respectivas grabaciones y se generó los modelos requeridos para el sistema de reconocimiento del habla y posteriormente pueda generar el subtítulo correspondiente.

Dentro de este documento se explicará de manera detallada como está desarrollado el sistema, se desglosará su estructura y se detallará el funcionamiento del mismo, de modo que el lector pueda entender con exactitud el mismo y pueda ser de gran utilidad para el desarrollo de futuros proyectos.

2. OBJETIVOS

2.1. General

Diseñar y desarrollar una herramienta capaz de capturar el audio en lenguaje español (Ecuador) convirtiéndolo a texto en tiempo real, y ser usado como apoyo para personas de tercera edad y aquellas que presentan discapacidad auditiva.

2.2. Específicos

- Investigar y analizar los fundamentos básicos de los sistemas de reconocimientos de voz.
- Revisar y definir el algoritmo y la secuencia lógica para realizar el proceso de extracción de texto.
- Implementar el algoritmo y los procesos de reconocimiento automatizado de conversión de audio.
- Analizar, identificar e implementar el proceso óptimo para la de decodificación y encontrar la mejor secuencia de palabras emparejadas que coincidan con el discurso de entrada.
- Sincronizar y visualizar del texto resultante en el dispositivo escogido para su presentación.
- Diseñar y ejecutar un plan de experimentación que permita validar el sistema desarrollado.

3. ESTADO DEL ARTE

3.1. La discapacidad Auditiva a nivel Mundial

La Organización mundial de la salud nos revela las cifras del número total personas que padecen discapacidad auditiva a nivel mundial, el 5% de la población existente es decir 360 millones de personas presentan esta discapacidad [5], de los cuales 328 millones son adultos y 32 millones son niños.

Al hablar de discapacidad auditiva debemos tener en cuenta que dentro de esta discapacidad podemos distinguir 2 tipos:

- Sordera: La sordera abarca una pérdida auditiva severa y profunda las cuales que se encuentra en el rango de 70 y 90 dB y superior a los 90 dB, en este punto es imprescindible el uso de prótesis en ambos oídos para mejorar su audición [6].
- Hipoacusias: En el caso de la hipoacusia, esta abarca la deficiencia auditiva ligera y media las cuales se tienen una pérdida auditiva de entre 20 y 40 dB, 40 y 70 dB, estos problemas no afectan en gran magnitud la calidad del sonido, sino que afecta la cantidad del mismo [6].

3.1.1. Problemas asociados a la discapacidad auditiva

El correcto funcionamiento de los órganos sensoriales aporta una base fundamental en el adecuado desarrollo evolutivo de un individuo, por lo que el no escuchar con claridad o en el peor de los casos no escuchar, genera en las personas que lo padecen limitaciones a la hora de realizar las tareas en las diferentes áreas de su vida cotidiana, áreas como la educativa, social, familiar y laboral, son afectadas por el gran muro que limita e impide su acceso a la información, esto también puede desencadenar en el individuo síntomas como: ansiedad, depresión, baja autoestima, que puede llegar a afectar en gran medida su calidad de vida, en el ámbito laboral podrían llegar a perder su empleo [7].

3.1.2. La discapacidad auditiva en el Ecuador

El Ministerio de Educación del Ecuador, define a la discapacidad auditiva como: “la limitación sensorial que presenta una persona en relación al sentido de la audición que puede afectar a uno o ambos oídos (unilateral o bilateral); esta pérdida sensorial es ocasionada porque los órganos y estructuras que permiten la percepción de los estímulos sonoros se encuentra afectada” [8].

Según los datos recopilados por el Consejo Nacional para la Igualdad de Discapacidades, en el Ecuador existen un total de 447.294 personas con discapacidad registradas [9], de los cuales un total de 63.191 de las personas

presentan discapacidad auditiva, siendo así la tercera discapacidad que más afecta a la población Ecuatoriana.

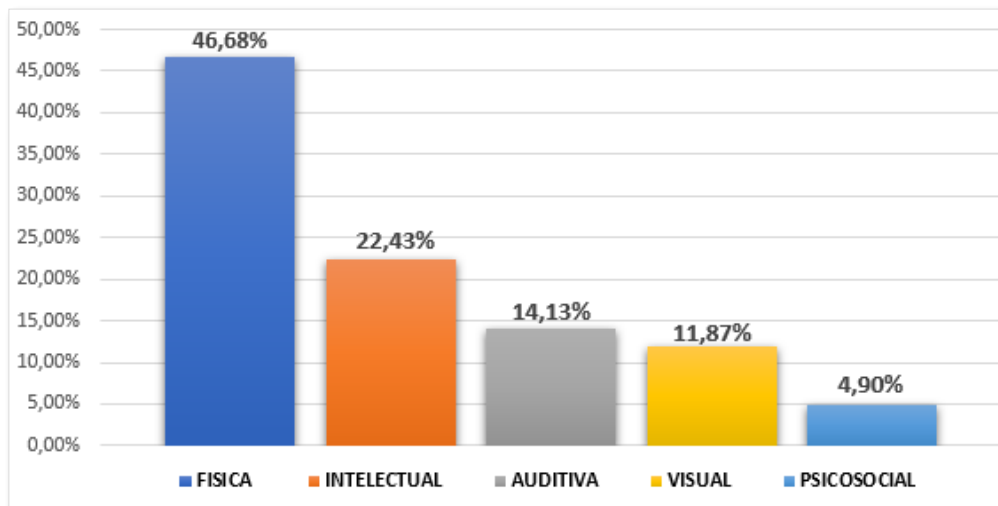


Figura 1. Tipos de Discapacidad en Ecuador.

En cuanto al grado de discapacidad podemos observar que en la figura 2, que el mayor grupo de personas con esta discapacidad posee un grado de discapacidad de 30% a 49% [9], es decir que tiene dificultades para escuchar con claridad pero que no han perdido por completo este sentido de audición.

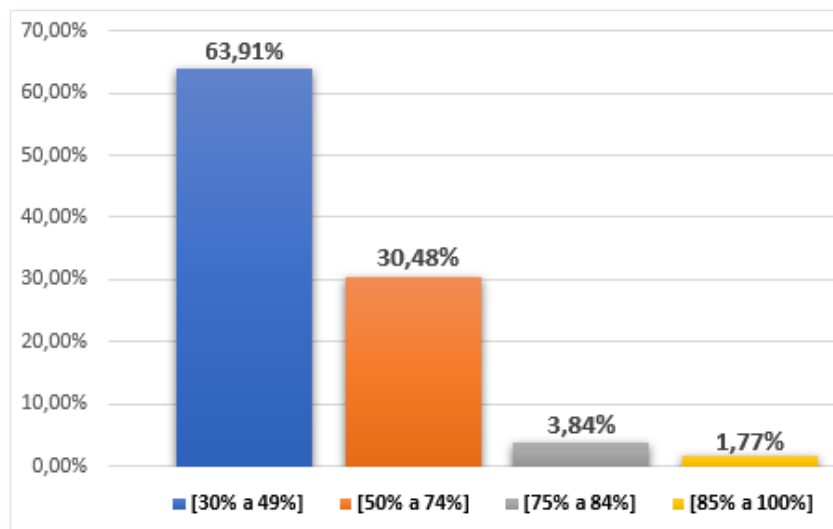


Figura 2. Discapacidad en Ecuador, grados de discapacidad auditiva

En cuanto a los grupos etarios, podemos notar claramente en la figura 3 que el grupo mayormente afectado por esta discapacidad son los adultos con edades mayores 30 años, seguidos por los adultos mayores con edades mayores a 65 años [9].

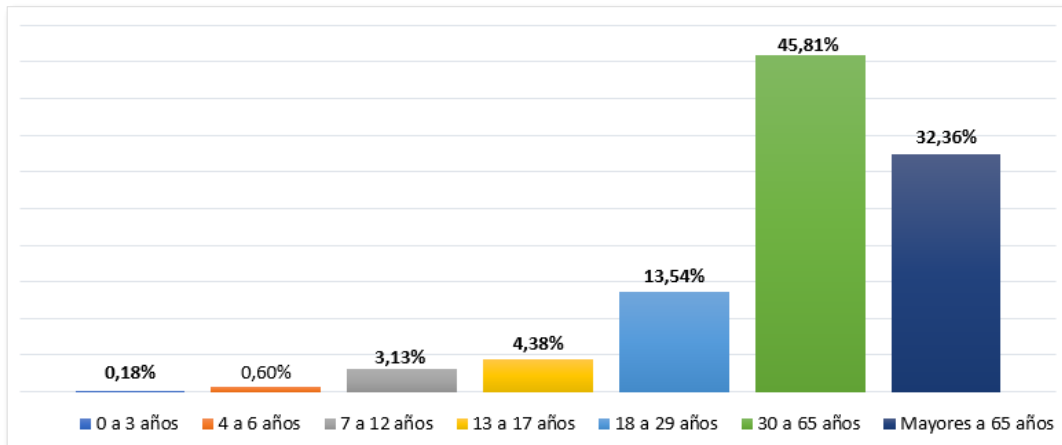


Figura 3. Discapacidad en Ecuador, grupos etarios

3.2. El Adulto Mayor

Según las Naciones Unidas, adulto mayor o anciano a aquella persona que sea mayor a 65 años para los países desarrollados y en el caso de los países en desarrollo son 60 años. La Organización Mundial de la Salud nos dice también que las personas que tienen edades a partir de los 60 a 74 años son consideradas de edad avanzada [10], las personas que se encuentran entre 75 a 90 años son denominadas ancianas y las que están por sobre los 90 años se les considera grandes longevos. Personas de la tercera edad, es el término utilizado para referirse a aquellas personas cuya edad sobrepasa a los 60 años.

3.2.1. Adulto Mayor y Discapacidad

Debemos tener en cuenta que la vejez es una etapa más de la vida de la que nadie está exento, pues este es un proceso en el cual las funciones cognitivas y sensoriales se encuentran en un deterioro progresivo. El ritmo de este proceso varía en cada persona, los problemas o síntomas que se ven reflejados en la tercera edad, son las repercusiones de los comportamientos y la calidad de vida que llevamos durante nuestra juventud.

Según un estudio realizado por la Organización Mundial de Salud afirma que “En los años futuros, la discapacidad será un motivo de preocupación aún mayor, pues su prevalencia está aumentando. Ello se debe a que la población está envejeciendo y el riesgo de discapacidad es superior entre los adultos mayores, y también al aumento mundial de enfermedades crónicas tales como la diabetes, las enfermedades cardiovasculares, el cáncer y los trastornos de la salud mental.” [11] .

Este grupo etario se enfrenta al desgaste de los sentidos y en el peor caso una pérdida total del mismo, pues a su edad tanto los problemas auditivos como los problemas oftalmológicos son prácticamente inevitables, los problemas de visión se presentan en un inicio como una molestia al tener que alejar los objetos de lectura

para poder enfocarlos y leerlos sin molestia, visión borrosa, dolor en los ojos, manchas en su campo de visión, pérdida visual ya sea a nivel lateral o periférica, etc.

Cada uno de estos síntomas puede llegar a desembocar en enfermedades asociadas con la edad como cataratas, glaucoma, presbiacusia, glaucoma; mismas que no se pueden revertir, pero pueden ser tratadas con la finalidad de conservar el nivel de visión que todavía posee.

3.2.2. Adulto Mayor en Ecuador

Según las proyecciones realizadas por el Instituto Ecuatoriano de Estadísticas y Censos, para el año 2030, en la figura 4 podemos apreciar como Ecuador refleja una tendencia de reducción en el campo de natalidad, esto quiere decir que llegado el momento no nacerá suficiente gente que pueda reemplazar la población actual, de modo que habrá más personas que envejezcan que las que nacen.

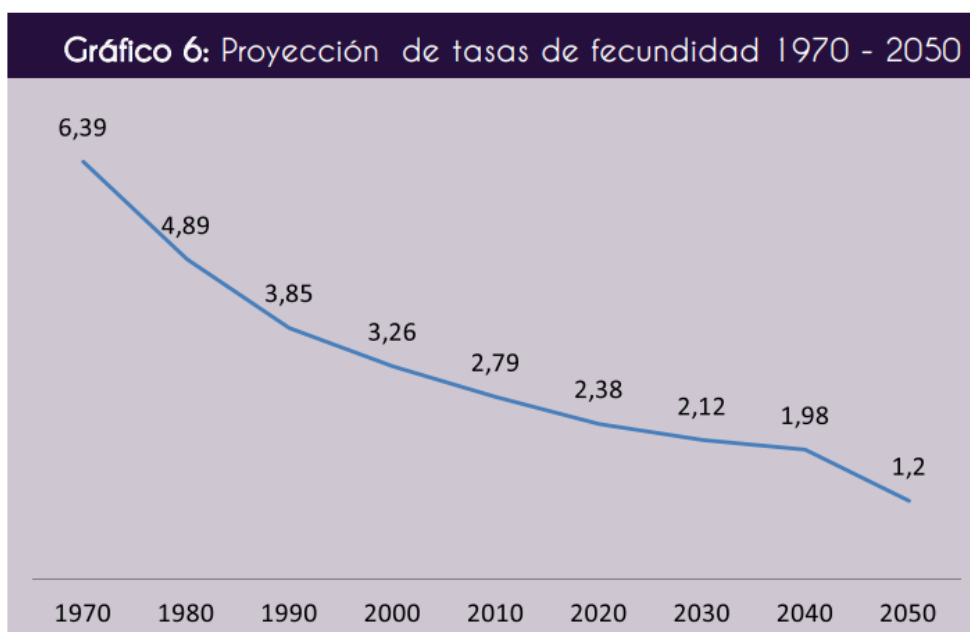


Figura 4. Proyección de Tasas de Fecundidad 1970-2050 [1].

3.3. Accesibilidad

ISO/TC 16027 define a la accesibilidad como “Facilidad de uso de forma eficiente, eficaz y satisfactoria de un producto, servicio, entorno o instrumento por personas que poseen diferentes capacidades. Accesibilidad electrónica hace referencia a que los productos y servicios electrónicos puedan ser utilizados por los usuarios con efectividad, eficiencia y satisfacción en un contexto de uso determinado” [12].

La accesibilidad en el mundo informático se presenta en soluciones tanto a nivel de hardware como a nivel de Software con la finalidad de garantizar y facilitar el acceso a los diversos recursos tecnológicos sin discriminar a la diversidad de

usuarios que haga uso del mismo, convirtiéndose así en un factor de apoyo que permita participar activamente en la sociedad y hacer uso de las nuevas tecnologías.

3.3.1. Accesibilidad para personas con discapacidad auditiva post-lingüística - adultos mayores

Como se mencionó anteriormente el adulto mayor se enfrenta a un desgaste progresivo de sus capacidades físicas y cognitivas, el mismo que varía en cada individuo, uno de los problemas a los que nos enfrentamos en la vejez es la pérdida gradual de la audición que limita el acceso a la información y por ello imposibilita una correcta participación.

Al darse la discapacidad auditiva en una fase de desarrollo de lenguaje post lingüística o de forma gradual, entendemos que esta persona tiene conocimiento del lenguaje oral y del lenguaje escrito por lo que pueden acceder a la información mediante la lectura de texto sin la intervención de un intérprete del lenguaje de señas, es por ello que a continuación se tratara los temas de accesibilidad en texto como un apoyo para mejorar su acceso a la información [6].

3.3.2. Subtitulado como apoyo para las personas con discapacidad auditiva

La principal herramienta como un medio de accesibilidad utilizada por las personas con discapacidad sensorial es el subtitulado, la cual es una transcripción en texto a partir de un discurso oral, con la finalidad de que todos puedan entender lo que se está presentando en ese momento [13].

La elaboración de contenido accesible debe asegurar un alto nivel de legibilidad para todos aquellos que accedan a este, para ello se debe hacer comprensible tres características: estructura, contenido y formato [14].

3.3.2.1. Pautas de accesibilidad para subtitulado

Es recomendable usar la Normativa UNE 153010 de “Subtitulado para personas sordas y personas con discapacidad auditiva”, como una guía a la hora de generar nuestros subtítulos. Cabe destacar que existen varias pautas que debemos tomar en consideración a la hora hacer de nuestro subtitulado accesible [13].

- **La combinación de colores entre el color de fondo y el color de fuente:**
Existen varias combinaciones de colores frecuente mente utilizadas que esta normativa nos ofrece entre ellas tenemos:
 - Fondo negro, caracteres amarillos
 - Fondo negro, caracteres verdes
 - Fondo negro, caracteres cian

- Fondo negro, caracteres magenta
- Fondo negro, caracteres blancos
- Fondo blanco, caracteres rojos
- Fondo blanco, caracteres azules
- Fondo amarillo, caracteres azules

Es importante que el color de fuente y fondo utilizados sean cómodos para el usuario, de modo que no provoque molestias en nuestra visión [13].

- **Tamaño de caracteres y el número de caracteres que se van a presentar por línea:**
Según esta normativa el número de caracteres recomendable es de 37 caracteres por línea, y el tamaño máximo de caracteres debe ser el que permita este número de caracteres. Es recomendable utilizar el tipo de letra más simple, es decir que no tengan decoraciones que dificulten la lectura
- **Interlineado entre textos:** El interlineado debe ser sencillo.
- **Número de líneas máximas y su ubicación:** Debe ocupar como máximo dos líneas y solo en casos especiales 3. El texto debe estar centrado
- **Tiempo de exposición en pantalla:** Es recomendable exponer 15 caracteres por segundo.

3.4. Fundamentos básicos de los Sistemas de reconocimiento de VOZ

Un sistema de reconocimiento del habla es una herramienta informática capaz de procesar las señales de voz emitida por el ser humano y reconocer la información contenida en ésta, y transcribiéndola en un texto legible. Para el desarrollo de este sistema es necesario tener en cuenta las siguientes disciplinas como: fisiología, acústica, lingüística, el procesamiento de señales, la inteligencia artificial y la ciencia de la computación [15].

3.4.1. Arquitectura de un sistema de reconocimiento de habla

La función principal de estos sistemas es capturar y analizar la señal de voz emitida, extraer la información acústica que contiene para finalmente realizar la representación en texto del conjunto de sonidos receptados. Para hacer posible este proceso de decodificación existen varios métodos que permiten analizar y comparar los diferentes patrones extraídos y devolver el resultado esperado que serán descritos más adelante, a continuación, enlistaremos y describiremos las fases que componen este sistema.

3.4.1.1. Pre procesamiento del audio

Dentro de esta fase, el sistema se encarga de realizar el proceso de captura y análisis de las señales de audio capturadas, para ello se realizan los siguientes procesos:

- **Extracción de características:**

Este proceso se encarga de extraer las características contenidas en la señal de voz que se está analizando, convirtiéndolos en una secuencia de vectores de características, para ello se hace uso de las Transformadas Rápidas de Fourier, se desechan también frecuencias innecesarias mediante un banco de filtro en la escala de Mel, posteriormente se aplica un logaritmo y la transformada discreta del coseno, teniendo como resultado un vector de coeficientes Cepstrales de la frecuencia de Mel [16].

- **Características robustas del ruido:**

Uno de los factores que afecta el proceso de reconocimiento es el ruido del ambiente, por ello es necesario remover el ruido, para ello existe dos algoritmos que pueden ser implementados en este proceso, el algoritmo SPLICE (Stereo-based piecewise linear compensation for environments) y el algoritmo QE (Quantile-based histogram equalization) [16].

- **Estimación de características:**

La variación en las características acústicas pueden ser debido a diversos locutores o en el mismo locutor, ocasionando que dificulte el proceso de reconocimiento por lo que es necesario eliminar estas variaciones, para ello se debe generar características adaptativas y discriminativas, las técnicas existentes para este proceso son normalización de la longitud del tracto vocal (VTLN), transformación de las características maximizando la verosimilitud bajo un modelo actual (fMLLR) [16].

3.4.1.2. Reconocimiento de Patrones

En este proceso se debe implementar uno de los métodos descritos más en la sección 5.4.2, estos métodos son implementados para obtener una representación de los patrones, más conocida como modelo matemático, misma que son utilizadas como referencia durante las comparaciones de patrones. Este proceso está constituido por dos subprocesos que se describirán a continuación:

- **Entrenamiento:**

Consiste en la creación de un patrón de referencia directamente relacionado a cada palabra o unidad fonética que se desea reconocer, esto lo hace posible mediante el uso de sus vectores de características

correspondientes, dado como resultado un modelo acústico [16]. Tenemos varias formas de entrenar a nuestro sistema y las describiremos a continuación:

- Entrenamiento casual: Asigna un solo patrón de sonido para la generación de modelo estadístico aproximado.
- Entrenamiento robusto: En este caso se asignan varias versiones de las unidades de palabras que desea reconocer, generando de este modo un modelo estadístico promedio.
- Entrenamiento por clustering: Busca generar modelos estadísticos con un mayor nivel de confianza y acierto por lo que requiere de grandes cantidades de datos, de manera que tenemos varias versiones de cada unidad de palabra a reconocer [16].

- **Comparación:**

En esta última etapa se realiza una comparación entre el vector de características generado a partir de las señales de voz y los todos los patrones obtenidos en la fase de entrenamiento con el único objetivo de encontrar el que mejor se ajuste a la señal que se pretende reconocer. Para ello establece una medida de similaridad entre los vectores antes mencionados, generando un modelo de lenguaje [16].

3.4.1.3. Decisión

Como es la fase final, es la encargada de tomar la decisión absoluta sobre la designación de los patrones que se han generado en la fase de entrenamiento, para ello hace uso de las medidas de similaridad establecidas en la fase de comparación, toma este valor y según su mayor nivel de similitud entre el modelo acústico y el modelo de lenguaje procede a tomar su decisión y presentar el resultado obtenido al usuario [16].

3.4.2. Métodos que emplean

- **Modelos ocultos de Márkov (HMM)**

Es una técnica probabilística utilizada para modelar las distribuciones de probabilidad, es decir que su finalidad es determinar los estados desconocidos partiendo de la secuencia de acciones que se genera de los valores observables, es por ellos que son implementadas con gran éxito para el desarrollo de sistemas de reconocimiento automático del habla.

Un HMM "es una máquina de estados finitos capaz de generar secuencias de observaciones que son asociadas con alguna unidad del habla" [17], es entrenado con la finalidad de que represente de la mejor forma posible las secuencia para una unidad determinada que en este caso pueden ser fonemas, palabras y silabas [18].

A continuación, podemos apreciar la arquitectura de este tipo de modelos, las flechas continuas hacen referencia a las posibles transiciones entre estados que se puede dar, en el primer caso de la figura 5, las transiciones se dan solamente desde la izquierda hacia la derecha mientras que en el segundo caso puede partir desde cualquier estado haciendo posible todas las combinaciones de transiciones [18]:

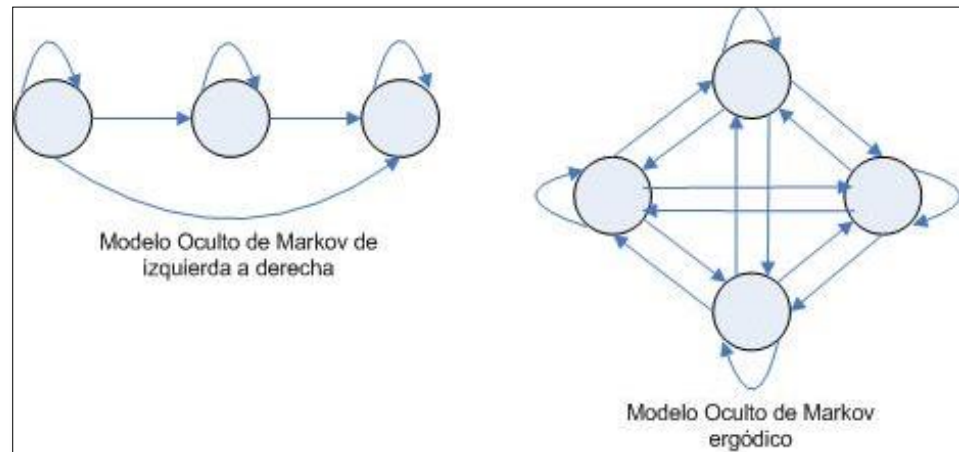


Figura 5. Arquitectura de los modelos ocultos de Márkov.

- **Redes neuronales artificiales (RNA)**

Las redes neuronales artificiales están inspiradas y creadas de manera que puedan imitar el funcionamiento del cerebro, con la finalidad de resolver problemas mediante el uso del conocimiento que han adquirido en base a casos similares.

Una red neuronal está compuesta por un conjunto de neuronas, en la siguiente figura podemos observar cómo está constituida la neurona, podemos observar que los valores de X_n son los valores de entrada, W_{ij} son los pesos sinápticos, $f()$ es la función de activación o transferencia y X_i es el resultado o salida de la neurona [19].

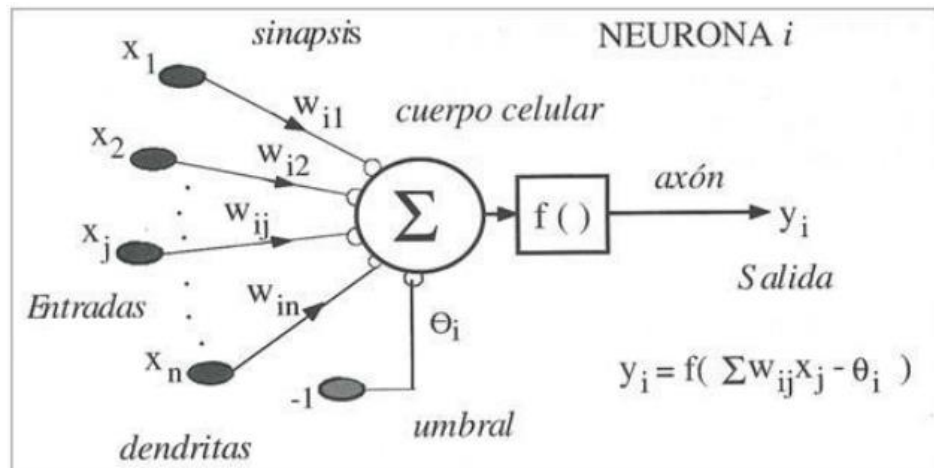


Figura 6: Modelo de una neurona artificial o Perceptrón simple [19].

El modelo de red neuronal más utilizado en el campo de reconocimiento del habla es la red de kohonen, ya que tiene la peculiaridad de tener de forma organizada sus pesos, mismo que son ajustados en función de la neurona que obtuvo el resultado, de este modo si posteriormente se presenta un patrón similar esta neurona responderá con mayor fuerza [20].

3.4.3. Modelos que emplea un sistema de reconocimiento del habla

Durante la fase de reconocimiento de patrones se generan el modelo acústico y el modelo de lenguaje en base a la aplicación los diferentes métodos antes mencionados, en esta sección se describirá la función que desempeña cada uno. Los recursos necesarios que un sistema de reconocimiento del habla utiliza para llevar a cabo esta tarea son los siguientes:

- **Modelo acústico:**

El modelo acústico contiene las propiedades acústicas para cada unidad sub fonética, es decir que representa el habla y/o la pronunciación emitida por el usuario de tal manera que el sistema pueda entender y posteriormente procesar [21]. A continuación, podemos observar como las señales de audio son segmentadas y procesadas para posteriormente ser representada mediante vectores de características.

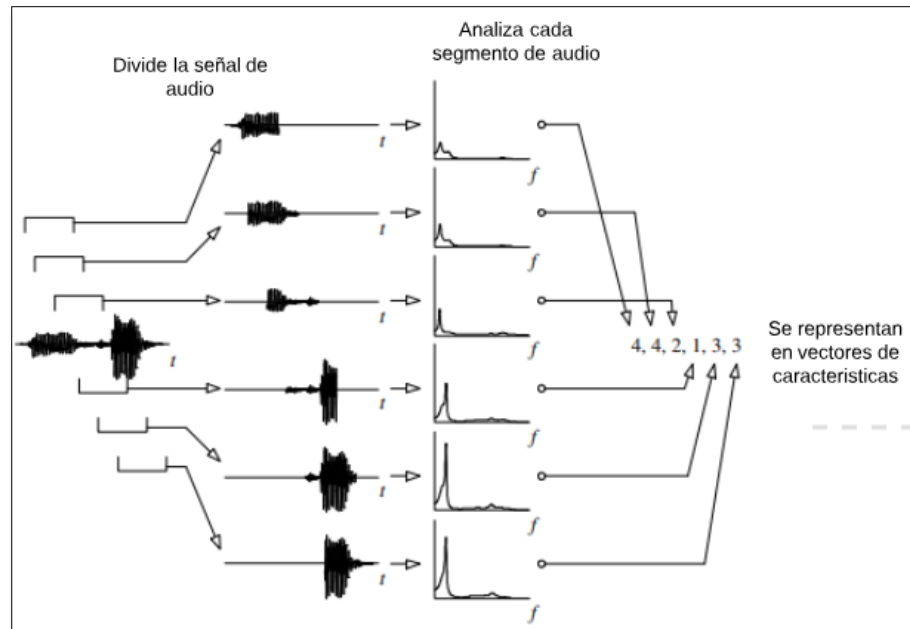


Figura 7. Procesamiento de señales para estructurar el modelo acústico [18].

- **Diccionario fonético:**

Proporciona al sistema un mapeo de todas las palabras y de cada una de ellas su respectiva secuencia de fonemas. La principal razón para la creación de este diccionario es que en ocasiones las palabras pueden tener una o más pronunciaci3nes alternativas, y todas ellas son agregadas al diccionario. El diccionario debe contener todas las palabras que el sistema va a reconocer. [21]

```

a a
aaron a a r o n
ab a b
abajo a b a j o
abandona a b a n d o n a
abandonada a b a n d o n a d a
abandonadas a b a n d o n a d a s
abandonado a b a n d o n a d o
abandonados a b a n d o n a d o s

```

Figura 8. Estructura del diccionario fonético espa3ol Ecuador.

- **Modelo de lenguaje:**

Este es un modelo estadístico que contiene las secuencias de palabras y la probabilidad que le corresponde a cada una, de esta forma se encarga de comunicarle al decodificador cuales son las secuencias de palabras que son posibles de reconocer y define qu3 palabra podría seguir a la secuencia de palabras previamente reconocidas.

Es importante tener en cuenta que el sistema reconocedor busca la palabra en el diccionario fonético y en el modelo del lenguaje, y en el

caso de que aquella palabra no se halle en el modelo del lenguaje el sistema lo omite, de este modo reduce el espacio la búsqueda de palabras. [21]

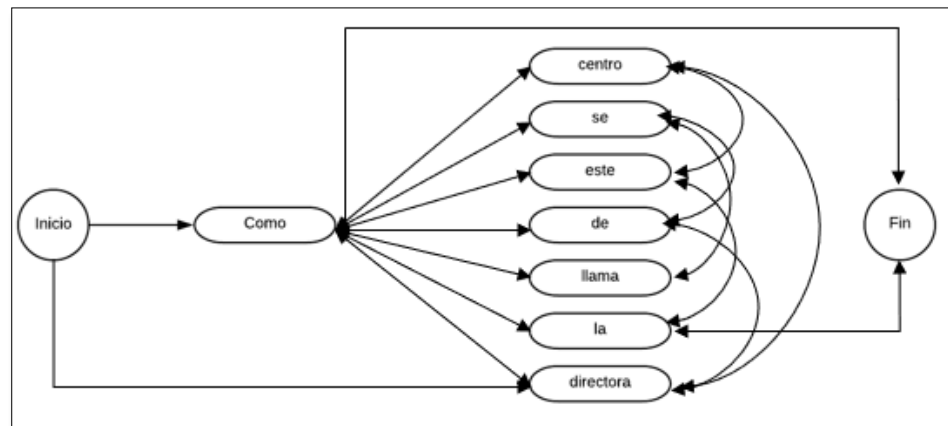


Figura 9. Estructura del modelo de lenguaje.

3.4.4. Herramientas existentes

- **CSLU- Toolkit**
es una librería de software que contiene varias herramientas que permiten crear un reconocedor basado en redes neuronales [22].
- **Api Speech de Google Cloud**
Api que permite la conversión de audio a texto mediante la implementación de redes neuronales para el proceso de reconocimiento, además de que permite realizar tareas como: transcripción de texto dictado por micrófono o a partir de un archivo de audio, control por voz entre otras [23].
- **JavaScript Web Speech API**
Esta API posee un gran nivel en las capacidades de reconocimiento de voz en aplicaciones web, pero hay que tener en cuenta que solo es posible en las versiones de Chrome igual o posteriores a la 25 [24].
- **CMUSphinx:**
Ofrece un conjunto herramientas para el desarrollo de aplicaciones orientadas al reconocimiento de voz, funciona de forma offline. Utiliza modelos matemáticos probabilísticos (HMM) para la representación del habla [25].
- **TensorFlow:**
Es una librería de código abierto cuya base central es el sistema de machine learning, especializado en realizar procesos de reconocimientos de voz, imágenes, textos, etc. [26]

- **PHNREC:**

Es un sistema que hace posible la detección de fonemas y ha sido aplicado en el campo de reconocimiento del habla, este sistema está actualmente entrenado y probado para idiomas como el checo, ruso, húngaro e inglés, siendo el primero mencionado el que más nivel de acierto posee [27].

3.5. Trabajos Relacionados

La implementación de del reconocimiento del habla en gran medida ha sido para generar subtítulos de forma automática con la finalidad de que estos medios audiovisuales sean accesibles para todo el público.

Del mismo modo en España Revuelta, Jiménez y otros [28] en su proyecto nos presentan las tecnologías desarrolladas con la finalidad de brindar accesibilidad a los medios audiovisuales para personas con discapacidad, este sistema tiene una estructura cliente servidor, en el cual el servidor es la máquina del docente y los clientes son los dispositivos portátiles de los alumnos, cuenta con los siguientes procesos para realizar el subtítulo, un subsistema de reconocimiento de voz, el programa principal, un subsistema de gestión de conexiones y la aplicación que ejecuta la conversión de texto a voz.

Este sistema funciona de la siguiente manera, para la parte del reconocimiento del habla el sistema hace uso de la herramienta Dragon Naturally Speaking; desde el servidor primero se verifica que exista comunicación con todos los equipos conectados y luego ejecuta esta herramienta y procede a realizar la conversión de audio a texto y el resultado obtenido se puede visualizar de manera general en el proyector que usa el docente o también existe la opción de que el alumno pueda acceder a ello en su dispositivo portátil.

En Paraguay [29] se desarrolló un sistema que permite el reconocimiento del habla de uno de los idiomas ancestrales hablados en este país que es el idioma guaraní, en este trabajo se plantea el desarrollo de un modelo acústico en base a un entrenamiento previo, para esta fase se desarrolló una aplicación web que permita que las personas interesadas puedan ingresar a la página, y accedan a las frases en guaraní que contiene las misma, graban las frases en guaraní con sus voces y proceden a almacenarla en dicha página.

Universidad Nacional de Asunción

Así también en España se ha desarrollado una aplicación móvil gratuita denominada ability connect, la cual se presenta como una herramienta de apoyo para las personas con discapacidad auditiva, baja visión y dislexia, la cual conecta varios dispositivos en tiempo real ya se mediante bluetooth o wifi, el anfitrión crea una sesión y en el campo de texto empieza a redactar la información que está recibiendo esta se visualizara en los dispositivos conectados, se puede utilizar el teclado del dispositivo y su sistema de reconocimiento de voz para transcribir su

texto equivalente, claro que esta función depende de la conexión del dispositivo a internet.

4. ARQUITECTURA GENERAL DE LA PROPUESTA PLANTEADA

4.1. Diseño y Desarrollo de la aplicación de escritorio.

En la siguiente sección se detallará la estructura del sistema, los procesos que emplean, las librerías necesarias para su funcionamiento, sin olvidar que se detallara los requerimientos necesarios para la aplicación.

4.1.1. Requerimientos para el Desarrollo del Sistema.

Para el desarrollo del sistema se tomó en cuenta que será empleado en un centro gerontológico para el proceso de estimulación cognitiva en adultos mayores que pueden o no presentar discapacidad auditiva, fue necesario la intervención de un profesional en el área, la psicóloga de este centro fue quien nos brindó los datos e información necesarios para armar nuestro corpus de datos y entrenar nuestra aplicación. También se realizó un estudio de los estándares de accesibilidad para la presentación en pantalla del texto obtenido, con la finalidad de que el usuario pueda modificar ciertas características de la interfaz y pueda ser apreciado de la mejor manera.

4.1.2. Diseño y construcción de la arquitectura del sistema.

El sistema está desarrollado como una aplicación de escritorio pues como se mencionó anteriormente no tiene la necesidad de conectarse a internet para su funcionamiento y puede ser utilizada desde cualquier lugar. Por lo cual utiliza modelos ya entrenados y almacenados en la aplicación. El proceso de transcripción lo hace al tomar los datos emitidos por el hablante en el micrófono del computador los procesa y continúa con la transcripción.

Para la parte del reconocimiento de del habla se utilizó CMUsphinx el cual ofrece un conjunto de herramientas que hacen posible el proceso de transcripción de audio a texto, mediante el uso de un modelo acústico, un modelo de lenguaje y un diccionario fonético. Contamos con un corpus de audios que se utilizaron para la generación los modelos y diccionario antes mencionado

Toda la aplicación está desarrollada en Python que es un lenguaje de programación de código abierto que muy poderoso y al ser multiparadigma, permite la programación ya sea orientada a objetos y proporciona librerías que facilitan el desarrollo de aplicaciones.

4.1.3. Especificaciones técnicas la aplicación desarrollada

En la siguiente sección se describirá de forma detallada los requerimientos necesarios para la ejecución y funcionamiento de la aplicación desarrollada.

4.1.3.1. Especificaciones de hardware

- **Procesador:** Intel Core I7
- **Memoria RAM:** 8 GB
- **Disco Duro:** 1 TB

4.1.3.2. Especificaciones de software

- **Sistema operativo:** Ubuntu 16.04

4.1.3.3. Herramientas de software elegidas para el desarrollo

Herramienta	Versión	Sitio de descarga
Ubuntu	16.04	https://www.ubuntu.com/download/desktop
Python	2.7	Gestor de paquetes Ubuntu
Numpy	1.15.4	http://www.numpy.org/
Matplotlib	2.2.3	https://matplotlib.org/
Struct	2.5.2	Gestor de paquetes Ubuntu para Python2.7
Wave	3.9.0	Gestor de paquetes Ubuntu para Python2.7
sounddevice	0.3.12	Gestor de paquetes Ubuntu para Python2.7
urllib3	1.24.1	Gestor de paquetes Ubuntu para Python2.7
QT Designer	4.0	Gestor de paquetes Ubuntu para Python2.7
PyQt4	4.11.4	Gestor de paquetes Ubuntu para Python2.7
Pyaudio	1.3.0	Gestor de paquetes Ubuntu para Python2.7
Pip	9.0.1	Gestor de paquetes Ubuntu para Python2.7
Pyqtgraph	0.10.0	Gestor de paquetes Ubuntu para Python2.7
PocketSphinx	5prealpha	Gestor de paquetes Ubuntu para Python2.7
SphinxBase	5prealpha	Gestor de paquetes Ubuntu para Python2.7
SphinxTrain	5prealpha	Gestor de paquetes Ubuntu para Python2.7

Tabla 1. Herramientas utilizadas, con su respectiva versión

4.1.3.4. Python

La página oficial de Python nos dice que es “un lenguaje de programación que le permite trabajar rápidamente e integrar sistemas de manera más efectiva” [30], al ser un lenguaje de programación orientado a objetos puede ser aplicado para el desarrollo de diversos programas y/o aplicaciones.

Se escogió Python como el lenguaje fundamental para el desarrollo de nuestra aplicación gracias a las librerías y paquetes con las que se pueden trabajar nos facilita realizar tareas sin tener que construirlos desde cero, otra ventaja que podemos apreciar es que debido a que es su sintaxis no es nada complicada sino por el contrario es sencilla clara y fácil de interpretar [30].

4.1.3.5. Numpy

“Es una extensión de Python, que le agrega mayor soporte para vectores y matrices, constituyendo una biblioteca de funciones matemáticas de alto nivel para operar con esos vectores o matrices” [31].

4.1.3.6. **MalplotLib**

“Matplotlib es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python y su extensión matemática NumPy” [32].

4.1.3.7. **CMUSphinx**

Este es un conjunto de herramientas que posibilitan el reconocimiento de voz, siendo una de las herramientas más utilizadas para el desarrollo de aplicaciones que implementen este proceso. Esto lo hace posible ya que posee un conjunto de paquetes destinados para diferentes tareas; entre los paquetes que ofrece tenemos [25]:

- **PocketSphinx:** Esta biblioteca permite emplear el proceso de reconocimiento de voz y transformación a texto está desarrollada en el lenguaje de programación C.
- **Sphinxbase:** En esta biblioteca de soporte es requerida por la biblioteca antes mencionada para realizar sus funciones.
- **Sphinxtrain:** Esta biblioteca tiene una de las funciones más importantes cuando se quiere aplicar la herramienta para un nuevo idioma, pues nos ofrece varias herramientas de entrenamiento del modelo acústico.

Para el desarrollo de nuestro sistema utilizamos PocketSphinx debido a su velocidad y portabilidad, el sistema será empleado para reconocer el idioma español (Ecuador) por lo que es necesario pasar por una fase de entrenamiento previo, por lo cual es imprescindible el uso Sphinxbase y Sphinxtrain [25].

4.1.3.8. **Qt designer**

Es una herramienta que nos facilita la creación y diseño de interfaces para nuestra aplicación a nivel de usuario, ofrece un entorno en el cual se pueden arrastrar los componentes según las funciones que necesitamos para nuestra interfaz, también nos permiten integrar con el código programado en python y así ejercer acciones sobre ellos [33].

4.1.3.9. **PYQT4**

Es la combinación entre el lenguaje de programación Python y la librería GUI QT que ofrecen un conjunto de herramientas que facilita interacción y manipulación de la interfaz desarrollada en Qt desde el código desarrollado en python. Es compatible con sistemas operativos como Windows, Linux y MacOS [34].

Los dos módulos aplicados para la interacción con los componentes de la interfaz desarrollados en Qt designer son los siguientes:

- **QtCore:**

“Contiene las funcionalidades principales que no tienen que ver con la interfaz gráfica. Se usa para trabajar con tiempo, archivos y directorios, varios tipos de datos, flujos, urls” [35].

- **QtGui:**

“Contiene los componentes gráficos y sus clases relacionadas” [35].

4.1.3.10. Paquetes y librerías adicionales

Para lograr la interacción entre los diferentes lenguajes empleados y además lograr funcionalidades extras, necesarias para el correcto funcionamiento del sistema, se emplearon librerías externas, que nos ayudaron en el trabajo.

4.1.3.11. Audio – pydub, pyaudio, sounddevice

“PyAudio proporciona enlaces de Python para PortAudio, la biblioteca de E / S de audio multiplataforma. Con PyAudio, puede usar Python fácilmente para reproducir y grabar audio en una variedad de plataformas, como GNU / Linux, Microsoft Windows y Apple Mac OS X / macOS” [36], spunddevice se utilizó para reconocer los dispositivos de audios existentes.

4.1.3.12. Struct

“Este módulo realiza conversiones entre los valores de Python y las estructuras de C representadas como cadenas de Python. Esto se puede usar para manejar datos binarios almacenados en archivos o conexiones de red, entre otras fuentes” [37].

4.1.3.13. Wave

Esta librería permite la lectura y escritura de archivos de audio lo cual nos ha permitido representar en una gráfica su espectro. También “el módulo wav proporciona una interfaz conveniente para el formato de sonido WAV. No es compatible con la compresión / descompresión, pero sí es compatible con mono / estéreo” [38].

4.1.3.14. Urllib3

Esta librería se usó para conectar nuestra aplicación una herramienta web LMT00L que permite actualizar el modelo de lenguaje, para ello permite realizar las funciones POST y GET.

4.2. Diagramas de Secuencia

En esta sección podemos apreciar los diagramas de secuencias que representan los procesos más relevantes de nuestro sistema:

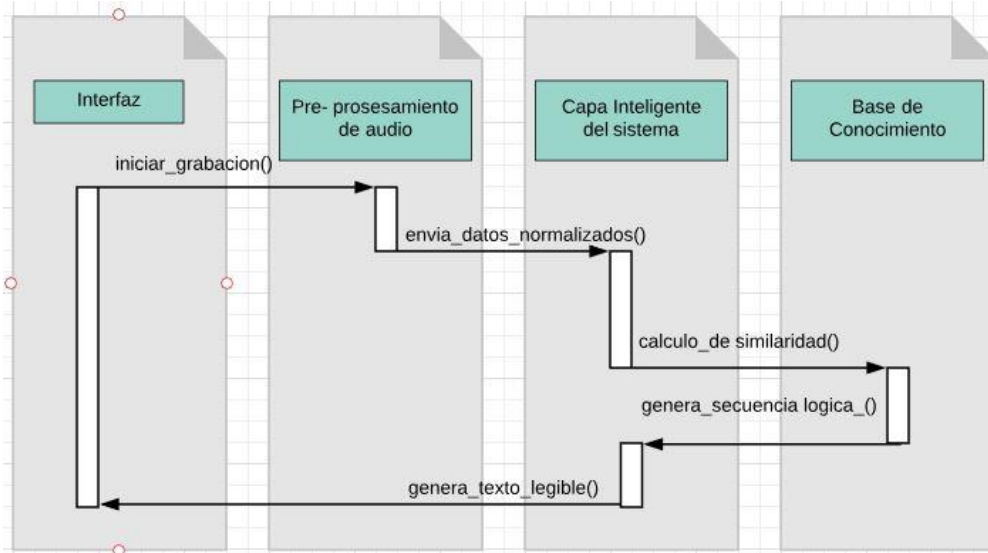


Figura 10. Diagrama de Secuencia: Reconocimiento y transcripción del habla

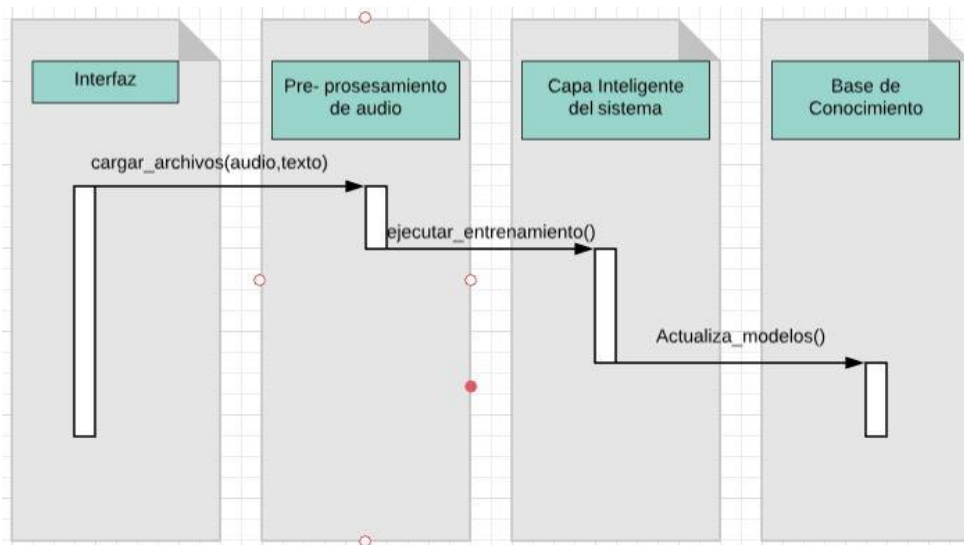


Figura 11. Diagrama de Secuencia: Ejecución del entrenamiento

4.3. Arquitectura general de la aplicación

La aplicación de escritorio desarrolla mantiene la siguiente estructura, primero está dividida en está dividida en 4 capas mismas que son: capa de interacción, preprocesamiento de auto, capa inteligente del sistema y capa de conocimientos, los cuales serán detallados más adelante, en conjunto cada uno de estos componentes hacen posible el correcto funcionamiento del sistema, más adelante se describe de manera detalla las funciones y procesos que cada capa desempeña.

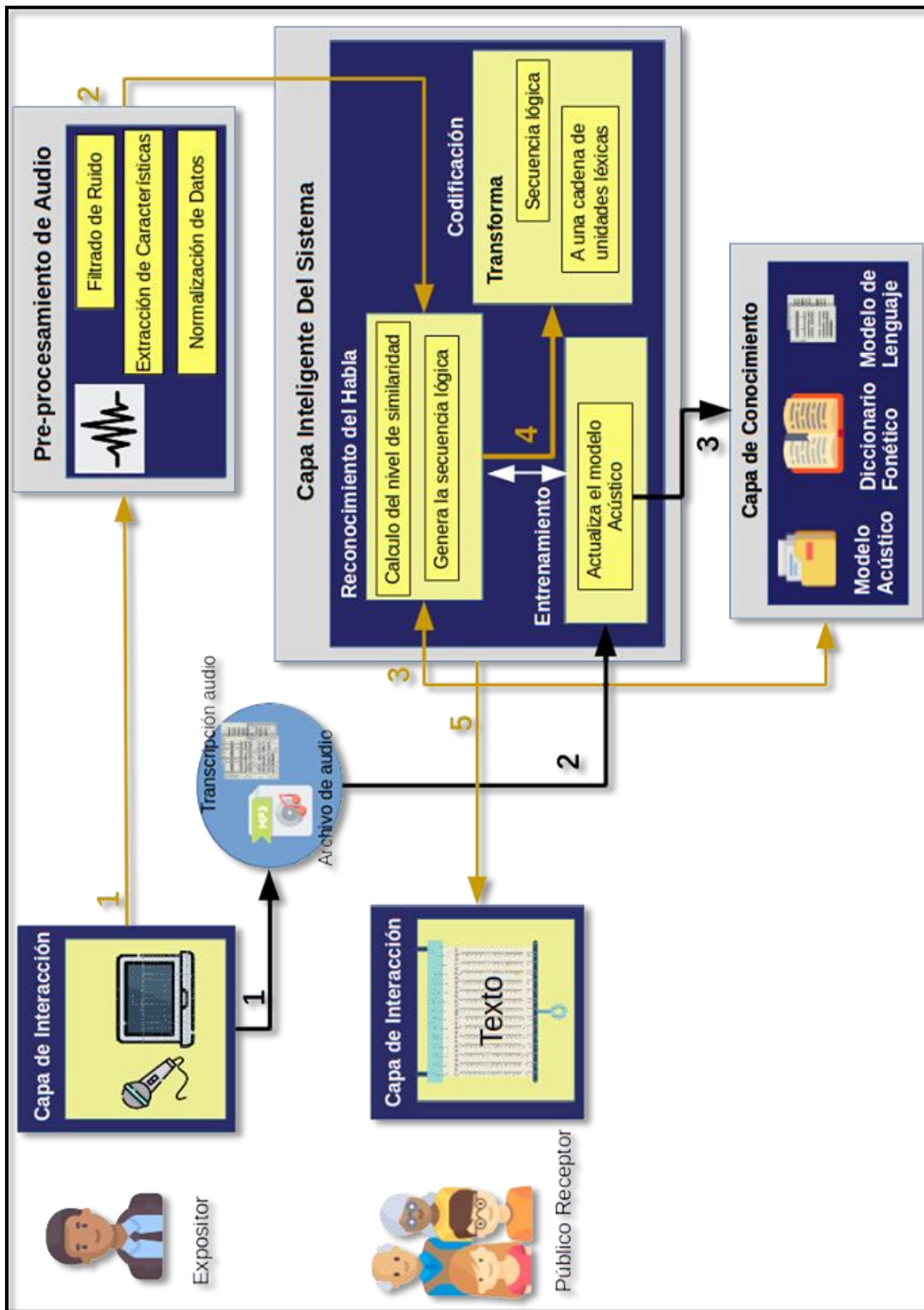


Figura 12. Arquitectura del Sistema

4.3.1. Capa de Interacción

Esta capa interactúa directamente con el usuario, y su función principal es la de recopilar las palabras emitidas por el usuario de manera secuencial, que serán

procesadas para extraer el texto, una vez realizado este proceso en las diferentes capas el texto obtenido será visualizado en la misma interfaz.

Dentro de esta capa tenemos los métodos necesarios para la captura de audio, para ello, primero se realiza un proceso de reconocimiento de los puertos de audio que posee el computador, posteriormente se verifica la presencia de un micrófono, se toma las ondas de voz que son receptadas por el mismo y se procede a grabarlos en un archivo .wav al mismo tiempo estos datos son representados en una sección de la interfaz como un espectro de voz que varía en relación al sonido hablado.

Después del procesamiento realizado en la capa de sistema inteligente en la base de conocimiento y el decodificador, tenemos como resultado el texto que representa al audio ingresado, este texto será presentado en pantalla en su respectiva sección, en donde también se puede modificar la combinación de colores de fondo y fuente respetando los estándares de accesibilidad para subtulado, también se puede variar el tamaño y ajustarlo según las preferencias del público receptor.

A continuación, podemos apreciar las interfaces correspondientes al sistema desarrollado:

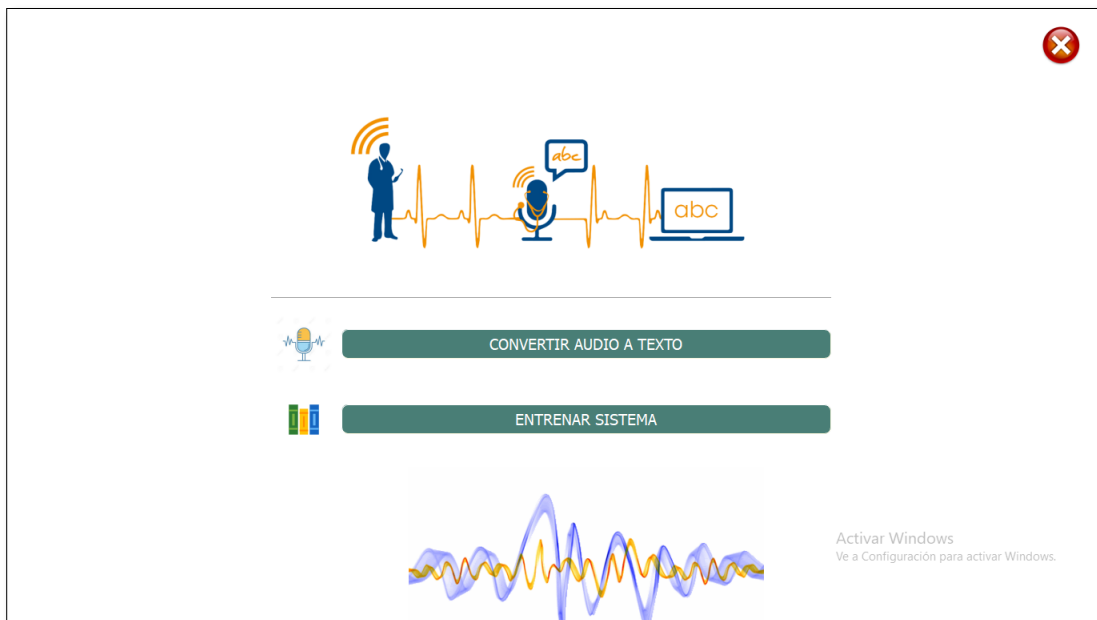


Figura 13. Pantalla Inicial de la Aplicación



Figura 14. Interfaz para el reconocimiento audio y transcripción a texto.

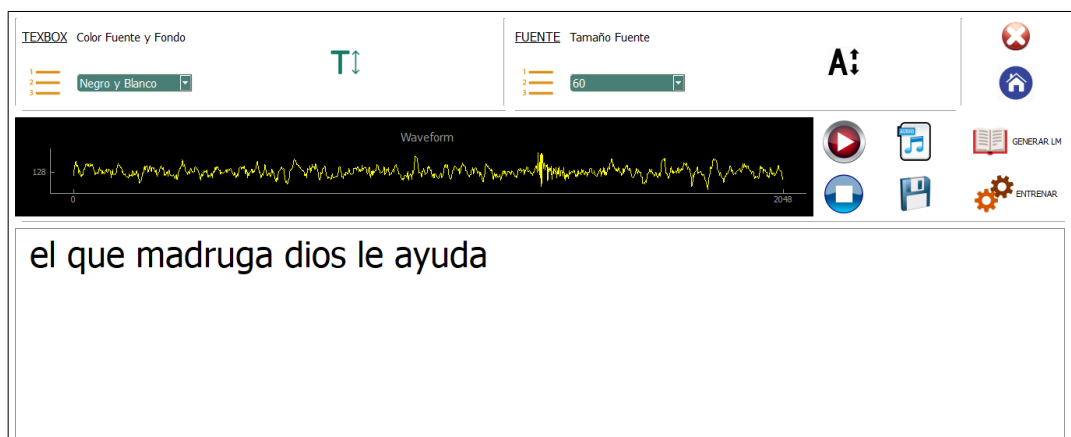


Figura 15. Interfaz para el entrenamiento del sistema.

4.3.2. Capa de Sistema Inteligente

En esta capa se implementan los procesos y la secuencia lógica que permiten el reconocimiento del habla y su posterior transcripción, el primer proceso que realiza es transformar el audio en un vector de características acústicas, con estos valores el sistema hace una búsqueda en la base de conocimiento, realiza una comparación y calcula la similitud entre el vector de entrada y los datos almacenados en los modelos, al finalizar este proceso, se envía la secuencia lógica de los fonemas al Decodificador.

Como sabemos que para cada fonema corresponde un estado del modelo oculto de Markov del modelo acústico, tenemos una secuencia de fonemas el resultado que nos envió la capa de sistema inteligente, pero este no es el resultado final, esto lo conseguimos mediante el uso del decodificador mismo que realiza la correspondencia entre la secuencia obtenida los datos que contiene el modelo de lenguaje, estructurando de esta manera un texto entendible para el usuario y es

retornado a la interfaz, para ello es necesario hacer uso de la herramienta PocketSphinx y establecer esta configuración para el análisis del audio de entrada.

```
config = {  
    'verbose': False,  
    'audio_file': os.path.join(data_path, 'goforward.raw'),  
    'buffer_size': 2048,  
    'no_search': False,  
    'full_utt': False,  
    'hmm': os.path.join(model_path, 'en-us'),  
    'lm': os.path.join(model_path, 'en-us.lm.bin'),  
    'dict': os.path.join(model_path, 'cmudict-en-us.dict')  
}
```

Figura 16. Configuración de los modelos utilizados por PocketSphinx.

4.3.3. Base de Conocimiento

En esta capa se encuentran el modelo acústico, modelo de lenguaje y el diccionario fonético que representan en conjunto el idioma español (Ecuador) mediante un conjunto de vectores de característica más probables (HMM), que posibilitan tomar todas las combinaciones posibles de palabras e intentar hacer coincidir las con el audio.

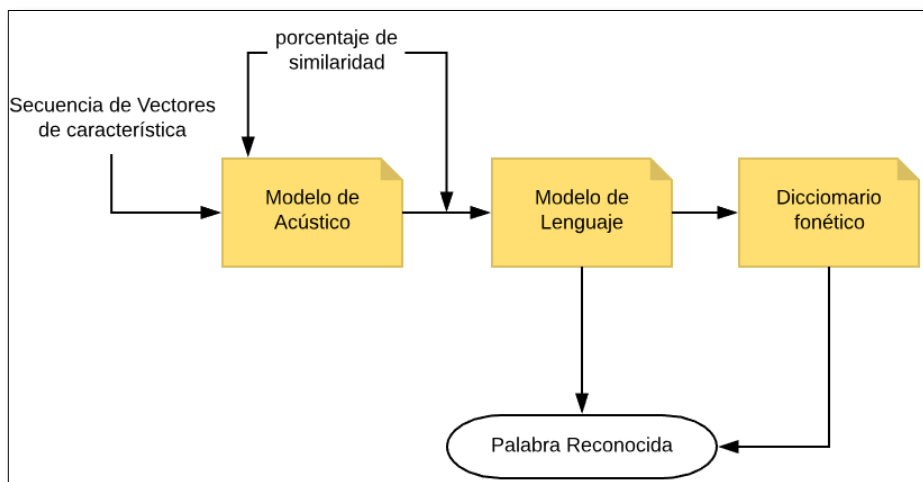


Figura 17. Funcionamiento e interacción de la base de conocimiento.

También se cuenta con el corpus de datos, está conformado por 350 audio, y los archivos necesarios para el entrenamiento, el primer archivo es el que contiene un listado con las etiquetas de los audios recopilados, el segundo archivo contiene las transcripciones correspondientes a todos los audios y está estructurado de la siguiente manera:

<s> transcripción del audio </s>(nombre_audio)

4.3.3.1. Entrenamiento y actualización del modelo Acústico y modelo de lenguaje

Para la ejecución del entrenamiento primero es necesario conocer los comandos que hacen posible los procesos que se describen más adelante:

a) **sphinx_fe**: herramienta de sphixbase que permite extraer las características de los archivos de audio también se debe especificar los siguientes argumentos:

- **-argfile**: se especifica la ruta del archivo que contiene los parámetros para la extracción de características, se encuentra dentro del modelo acústico como feat.params
- **-samprate**: se especifica la frecuencia de muestro de nuestros archivos de audio que es de 16000 hz.
- **-c**: Se especifica la ruta del archivo que contiene el listado de los nombres de los audios que en nuestro caso es (field.fileids).
- **-di**: se especifica la ruta de nuestro corpus
- **-ei**: se especifica el formato de los archivos de audio de entrada que deben ser en formato wav.
- **-eo**: se especifica el fromato que seran aplicaccdos a los datos de salida que sera mfc [25].

b) **Bw y map_adapt**: todas ellas son librerías ejecutables que nos proporciona sphinxtrain, para el proceso de adaptación del modelo acústico.

- **-hmmdir**: se especifica la ruta del modelo acústico.
- **-moddeffn**:
- **-dictfn**: se especifica la ruta del diccionario fonético.
- **-ctlfm**: Se especifica la ruta del archivo que contiene el listado de los nombres de los audios, en nuestro caso es (field.fileids).
- **-lsnfn**: Se especifica la ruta del archivo que contiene las transcripciones de los audios, en nuestro caso es (transcript.transcription).
- **-tmatfn**: se especifica la ruta del archivo las matrices de transición HMM del modelo acústico [25].

Para esta fase el usuario tiene la opción de ingresar a la interfaz que permite realizar el entrenamiento, una vez adentro debe cargar el archivo de audio que desea agregar al corpus posteriormente el sistema presenta el texto la transcripción, el mismo puede ser editado en el caso de presentar errores. Es importante que los audios no excedan los 30 segundos ya que es mejor para el sistema procesarlo, para entrenar al sistema se realizan los siguientes procesos:

- Primero extraemos las características de todos los audios, en este proceso de cada audio se genera un archivo que mediante vectores

numéricos o vectores de características representa las señales acústicas de cada audio para ello usamos el comando **sphinx_fe**.

```
os.system('sphinx_fe -argfile '+path_modelos+'/es-ec/feat.params \
-samprate 16000 -c '+path_training+'/field.fileids \
-di /home/ines/training_audios -do '+path_training+' \
-ei wav -eo mfc -mswav yes')
```

Figura 18. Comando para extraer las características de los audios.

- Durante la extracción de características el sistema se retroalimenta tanto del audio y la transcripción entregada con la finalidad de mejorar su algoritmo y mejorar la calidad de transcripción.
- El siguiente paso en la adaptación del modelo lo hacemos posible con las siguientes herramientas **bw**, **map_adapt**, primero se realiza una recopilación estadística de los datos que están siendo entrenados, luego se procede a actualizar con los datos obtenidos cada parámetro existente en el modelo acústico.

```
os.system('./bw \
-hmmmdir '+path_modelos+'/es-ec \
-modedefn '+path_modelos+'/es-ec/mdef \
-ts2cbfn .ptm. \
-feat ls_c_d_dd \
-svspec 0-12/13-25/26-38 \
-cmn current \
-agc none \
-dictfn '+path_modelos+'/cmudict-es-ec.dict \
-ctln field.fileids \
-lsfn transcript.transcription \
-accumdir .')
os.system('./map_adapt \
-modedefn '+path_modelos+'/es-ec/mdef \
-ts2cbfn .ptm. \
-meanfn '+path_modelos+'/es-ec/means \
-varfn '+path_modelos+'/es-ec/variances \
-mixwfn '+path_modelos+'/es-ec/mixture_weights \
-tmatfn '+path_modelos+'/es-ec/transition_matrices \
-accumdir . \
-mapmeanfn '+path_modelos+'es-ec-trainingFinal/means \
-mapvarfn '+path_modelos+'es-ec-trainingFinal/variances \
-mapmixwfn '+path_modelos+'es-ec-trainingFinal/mixture_weights \
-mapmatfn '+path_modelos+'es-ec-trainingFinal/transition_matrices')
os.system('pocketsphinx_batch \
-adcin yes \
-cepdir . \
-cepext .wav \
-ctl field.fileids \
-lm '+path_modelos+'/es-ec.lm.bin \
-dict '+path_modelos+'/cmudict-es-ec.dict \
-hyp adaptation-test.hyp')
```

Figura 19. Comando para la adaptación y actualización de los modelos.

Una vez actualizado el modelo acústico la aplicación genera un archivo que contiene las transcripciones de todos los audios en formato lm que mediante el uso de la herramienta lmtool que permite modelar y compilar el modelo de lenguaje, luego se procesa a actualizar los valores del modelo de lenguaje que se está usando en la aplicación.

5. EXPERIMENTACIÓN Y RESULTADOS

Según el ministerio de inclusión económica y social [39], nos presenta una norma técnica que regula el funcionamiento de los servicios de atención gerontológica mediante atención directa y/o por convenios.

Siendo el objetivo principal de los centros gerontológicos residenciales, mejorar la calidad de vida de sus habitantes mediante la restitución de sus derechos en base a una atención integral, que fortalezcan el goce de su bienestar físico, social y mental. Como uno de las obligaciones de un centro gerontológicos es brindar al adulto mayor de sesiones de estimulación cognitiva, pues estas ayudan a prevenir la pérdida de memoria y el desarrollo de enfermedades mentales como el Alzheimer. Para ello se emplean varios ejercicios que ayudan en el proceso de estimulación cognitiva, entre ellos tenemos el armar rompecabezas, la lectura, dibujar, memorizar refranes, clasificar objetos ya sea por forma o color, etc.

Es por ello que se para la evaluación de nuestro sistema que está orientado para mejorar la calidad de vida del adulto mayor como para las personas con discapacidad auditiva se llevó a cabo con la colaboración de la psicóloga del centro Gerontológico “Rosa Elvira de León” en la ciudad de Azogues, quien en primera instancia nos informo acerca de las sesiones de estimulación cognitiva que llevaba acabó con los residentes de esta institución, y en base a esto se procedió a estudiar las frases más utilizadas durante sus clases y enriquecer a nuestro sistema con ellas.

Posteriormente la evaluación de nuestro sistema se hizo posible mediante la aplicación de una encuesta la cual fue resuelta por los residentes de este centro, quienes con sus respuestas nos permitieron evaluar el nivel de acierto que la aplicación ofrece en su proceso de transcripción y constatar que la aplicación facilita y mejora el acceso a la información a personas con discapacidad auditiva y adultos mayores.

5.1. Pruebas

Para el proceso de experimentación en un inicio se planteó la posibilidad de mejorar los modelos del lenguaje español México ya existente mediante la implementación de una fase de entrenamiento para ello se tomó una muestra total de 345 audio, los cuales se obtuvieron de videos educativos que se encontraban en la plataforma web YOUTUBE. Para la primera fase de pruebas se utilizó el 20% de las grabaciones de nuestro corpus para realizar el test, los cuales son escogidos mediante un algoritmo que se encarga de leer todos nuestros archivos de audio los combina y los selecciona de manera aleatoria.

Los parámetros para esta evaluación son:

- El número total de palabras que reconoce el sistema
- El número total de palabras que contiene nuestro audio
- El número total de palabras que efectivamente el sistema reconoció con precisión

Al realizar las pruebas antes de la fase de entrenamiento se pudo observar que se presentaban casos en los que el sistema no lograba reconocer ni una palabra del audio de entrada, mismos que tiene un tiempo de duración entre 30 a 60 segundos, esto es debido a las peculiaridades y variaciones que existe tanto en el léxico como en la pronunciación y entonación que existe en los países de habla hispana.

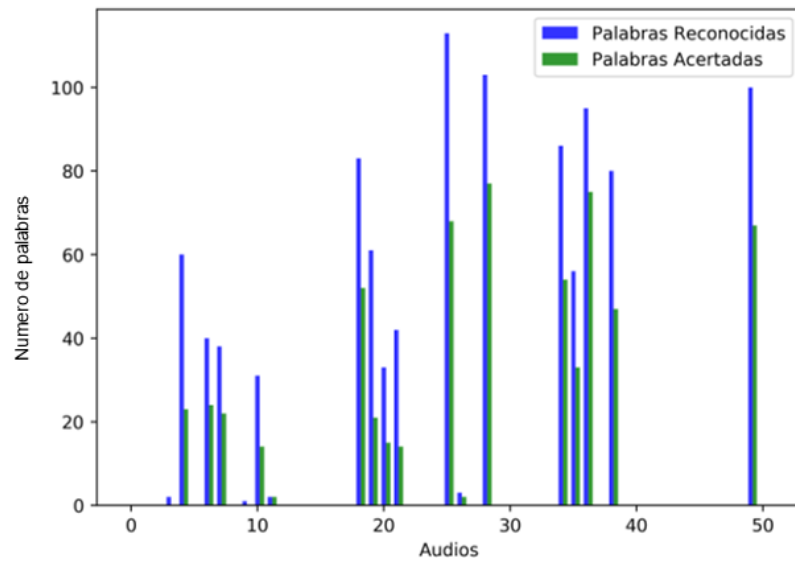


Figura 20. Pruebas del sistema de reconocimiento antes del entrenamiento del modelo acústico.

Al ejecutar las pruebas después de la ejecutar la fase de entrenamiento podemos apreciar en la Fig. 21 como el sistema mejoro notablemente pues ahora estamos usando un modelo acústico con las entonaciones, léxico y pronunciación propias de nuestro país.

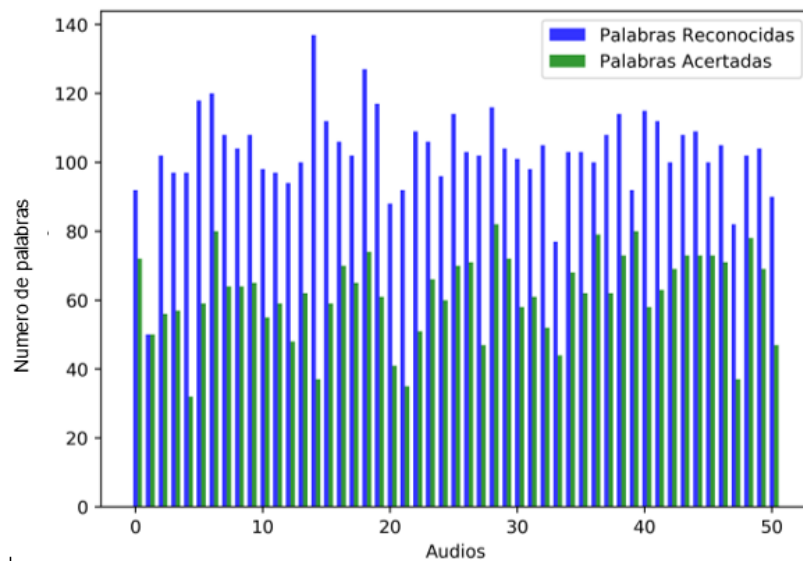


Figura 21. Pruebas del sistema de reconocimiento después del entrenamiento del modelo acústico.

Posteriormente al evaluar el nivel de comprensibilidad y claridad de las frases reconocidas al implementar la fase de entrenamiento en los modelos del lenguaje Español México se pudo llegar a la conclusión de que este proceso no era suficiente por lo que se procedió a generar los modelos que sean propio de nuestro país Ecuador.

Para proceder a evaluar la aplicación primero se recopiló un total de 45 frases utilizadas durante las clases de estimulación cognitiva misma que fueron grabadas por 2 personas, con esto se armó un corpus de 90 audios con el cual se procedió a entrenar previamente a la aplicación.

Para el análisis el nivel de funcionalidad se tomó una muestra de 35 frases de nuestro corpus las cuales están enfocadas en 4 áreas dentro del proceso de estimulación, a continuación, en la siguiente tabla se detallará las áreas en la que se trabajó y las frases por áreas.

Categoría	Objetivo	Frases
Orientación: Tiempo, espacio, persona	Mejorar el razonamiento lógico, ubicar situaciones en tiempo y espacio determinado	1. Qué día es hoy 2. En qué fecha estamos 3. Cómo se llama este centro 4. Cómo se llama la calle 5. Cómo se llama la directora de este centro 6. Qué queda al frente 7. Cómo se llama la ciudad 8. Como se llama la provincia 9. Como se llama el país
Lenguaje	Mejorar la capacidad comunicativa y expresiva.	1. Encado de conocerle 2. Es un placer saludarle 3. Cómo está usted 4. Hasta luego 5. Hasta Pronto 6. Hasta mañana 7. Estoy muy agradecido 8. Tengo hambre 9. Tengo calor 10. Me puede decir dónde queda esta calle
Memoria, Atención	Mejorar y mantener la capacidad de concentración, potencializar el aprendizaje, mantenimiento y recuperación de información	1. El que madruga Dios le ayuda 2. No hay mal que por bien no venga 3. Mas vale estar solo que mal acompañado 4. Caer está permitido, levantarse es obligatorio 5. A buen hambre no hay pan duro 6. A buen entendedor pocas palabras bastan
Funciones Ejecutivas	Mejorar la capacidad de iniciar, desarrollar y finalizar las acciones necesarias.	1. Ya está listo el refrigerio. 2. Ya le tengo que bañar. 3. Ya le tengo que cepillar los dientes. 4. Ya le tengo que cortar las uñas. 5. Ya le tengo que lavar las manos. 6. Venga a tomar café. 7. Vamos a misa. 8. Subamos a merendar. 9. Es hora del desayuno. 10. Es hora de visitar al doctor.

Tabla 1. Listado de frases con las que se trabajó en cada área.

Para iniciar las pruebas se aplicó una encuesta que previamente fue revisada y validada mediante el análisis de Cronbach con una puntuación de **0.870** según el cálculo del coeficiente alfa de Cronbach:

$$\alpha = \frac{k}{k-1} * \left[1 - \frac{\sum S_i^2}{S_T^2} \right]$$

Se pretende medir si la aplicación apoya su acceso a la comunicación y mejora su comunicación, se reunió a un total de 25 adultos mayores de los cuales se detalla sus datos y si presenta o no discapacidad auditiva en la siguiente tabla:

Nombre y Apellido	Edad	Nivel de Instrucción	Discapacidad auditiva	Tipo de sordera	Cuando ocurrió
Edgar Andrade	80	Primaria Incompleta	Si	Unilateral	Post-lingüística
Filomena Condo	85	Primaria Incompleta	No	Ninguna	Ninguna
María Torres	88	Primaria Completa	Si	Bilateral	Gradual
Luz Calle	62	Primaria Completa	No	Ninguna	Ninguna
Julio Vásquez	90	Primaria Incompleta	Si	Bilateral	Gradual
Dolores Jetón	75	Primaria Completa	No	Ninguna	Ninguna
Victoria Saldaña	96	Primaria Completa	No	Ninguna	Ninguna
David Tenenpaguay	75	Primaria Completa	Si	Bilateral	Gradual
Zoila Lema	93	Primaria Incompleta	Si	Unilateral	Gradual
Luz Álvarez	97	Primaria Incompleta	Si	Bilateral	Gradual
Ana Rodríguez	96	Primaria Completa	Si	Unilateral	Gradual
Margarita Jara	88	Primaria Incompleta	No	Ninguna	Ninguna
Manuel Aguilar	82	Primaria Incompleta	No	Ninguna	Ninguna
Ángel Montero	90	Primaria Incompleta	No	Ninguna	Ninguna
Victoria Cárdenas	67	Primaria Incompleta	No	Ninguna	Ninguna
Blanca Barrera	79	Primaria Incompleta	No	Ninguna	Ninguna
José Álvarez	96	Primaria Incompleta	No	Ninguna	Ninguna
Cesar Calle	69	Primaria Incompleta	No	Ninguna	Ninguna
Lastenia Flores	78	Primaria Incompleta	No	Ninguna	Ninguna
Ana Rea	79	Primaria Incompleta	No	Ninguna	Ninguna
Juana Torres	72	Primaria Incompleta	No	Ninguna	Ninguna
María Guamán	89	Primaria Incompleta	No	Ninguna	Ninguna
Jacinto Monzón	88	Primaria Incompleta	No	Ninguna	Ninguna
Manuel Orellana	60	Primaria Incompleta	No	Ninguna	Ninguna
Miguel Pacheco	85	Primaria Incompleta	No	Ninguna	Ninguna

Tabla 2. Datos de las personas encuestadas.

5.2. Resultados obtenidos

Las preguntas contenidas en la encuesta fueron calificadas en la escala de Likert, es decir, que se basa en una puntuación de 1 al 5, siendo 5 la mejor calificación y 1 la más baja calificación. Una vez finalizada la encuesta se analizaron los datos recopilados y se generaron los siguientes gráficos estadísticos:

5.2.1. Resultados de la evaluación de la aplicación en cada área

A continuación, podemos observar los porcentajes de acierto y claridad que el sistema ofrece en cuanto al proceso de reconocimiento y transcripción del habla,

para ello se realizaron pruebas de funcionalidad en base a la evaluación de cada frase detalladas en la tabla 1, correspondientes a cada área trabajada.

En la figura 22 podemos observar que los mejores resultados se dieron en el área de orientación, se puede apreciar que en esta área la aplicación tuvo un gran nivel de acierto en el subtítulo resultante con un porcentaje alentador de 82.22%.

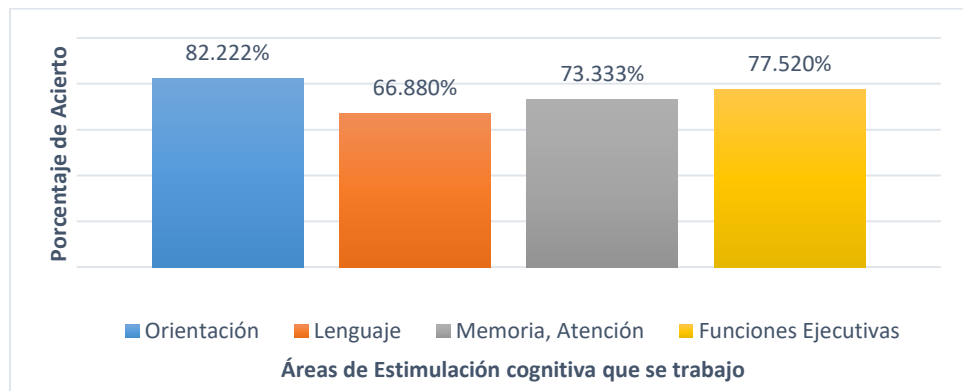


Figura 22. Nivel de acierto de la aplicación en las áreas de estimulación cognitiva trabajadas.

En la siguiente grafica estadística, se presenta un desglose de los resultados obtenidos en cada área. Es necesario explicar que los porcentajes de 20% indica que las frases no tenían coherencia y por ello nada entendibles.

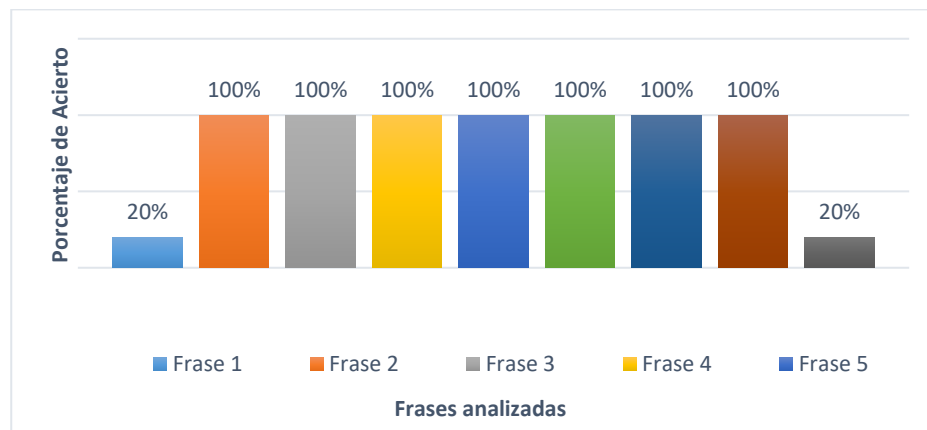


Figura 23. Nivel de acierto de la aplicación en el área de orientación

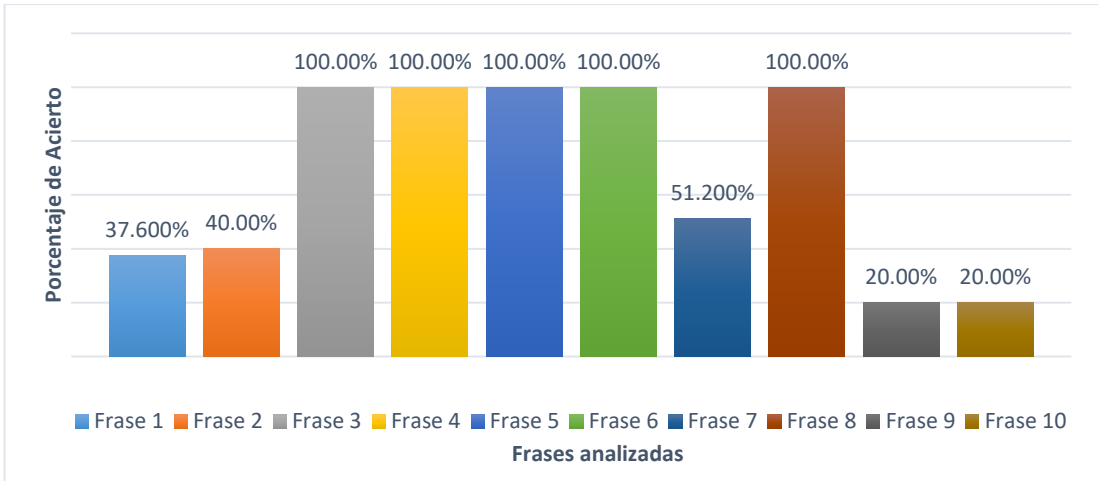


Figura 24. Nivel de acierto de la aplicación en el área de lenguaje.

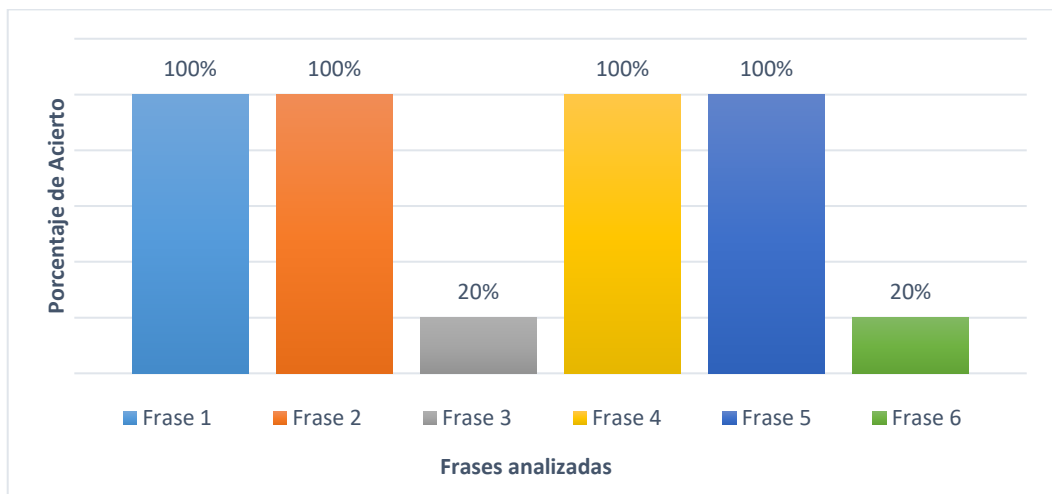


Figura 25. Nivel de acierto de la aplicación en el área de memoria y atención

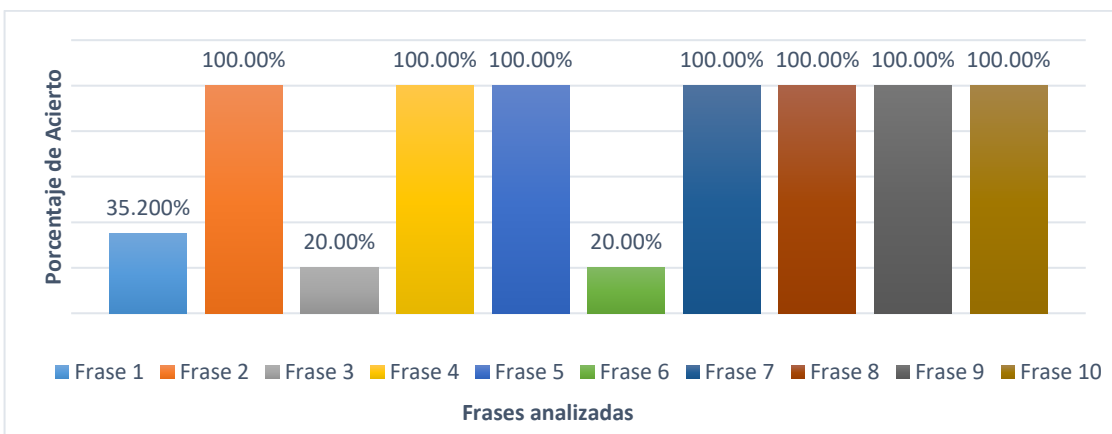


Figura 26. Nivel de acierto de la aplicación en el área de funciones ejecutivas

5.2.2. Resultados de la percepción del adulto mayor hacia la aplicación desarrollada

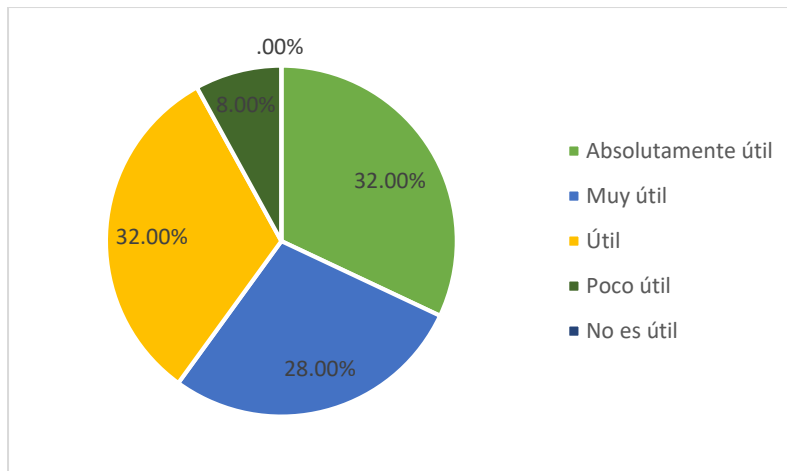


Figura 27. Estadísticas sobre utilidad de la aplicación.

5.2.3. Resultados del nivel de comprensión del adulto mayor después de implementar la aplicación en su terapia de estimulación cognitiva

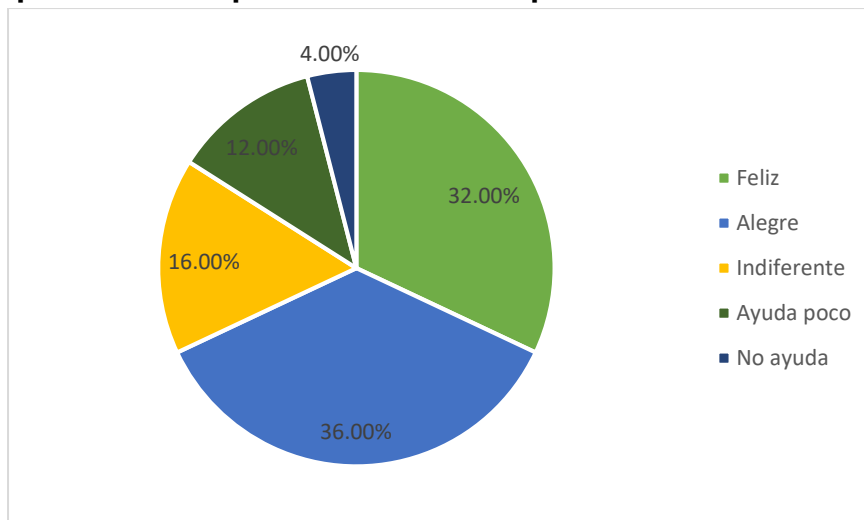


Figura 28. Estadísticas de implementación de la aplicación en las terapias.

6. CONCLUSIONES Y TRABAJO FUTURO

6.1. Conclusiones

Al iniciar trabajamos con modelos y diccionarios ya desarrollados para otras variantes del lenguaje español (México), pero como existe variaciones tanto en la pronunciación, entonación y léxico, el reconocimiento era muy pobre. En primera instancia se mejoró eso modelos mediante un proceso de entrenamiento. El sistema en un principio fue entrenado con audios de varias temáticas, con un total de 254 audios, pero presento varios problemas al realizar sus pruebas, si mejoro el nivel de reconocimiento, pero al no ser claras los textos presentados en pantalla podían a llegar a ser incoherentes y molestos en vez de ser un apoyo, es por ello que se optó por centrarnos en un solo tema e iniciar con un corpus pequeño, entrenarlo y generar los modelos para el lenguaje español(Ecuador) y como lo previsto los resultados fueron muy favorecedores.

Durante el desarrollo del presente proyecto hemos podido observar que la implementación de esta aplicación en el proceso de estimulación en el adulto mayor, mejoro el nivel de participación de quienes estaban presentes en este proceso, pues al no existir esa limitación de acceso a la información ellos comprendían las ordenes y/o preguntas que se le estaba haciendo y respondían con seguridad, en un principio pudimos observar que debido a su problemas de audición ellos no participaban activamente y no era porque desconocieran de las respuestas sino que por qué las escuchaban a medias y por ende no eran entendibles o no las escuchaban en su totalidad.

El haber convivido con los adultos mayores residentes en el centro gerontológico, me ayudo a conocer más acerca de los problemas que enfrenta el adulto mayor, pues al llegar a la vejez nuestras habilidades cognitivas se van deteriorando y poco a poco se va perdiendo la capacidad de recordar o pensar con lucidez por lo que las actividades las actividades desarrolladas en estimulación cognitiva los ayuda a mejorar y/o sobrellevar esta realidad. Es por ello que fue indispensable la participación de la psicóloga del centro gerontológico pues ella fue quien nos brindó sus conocimientos en el área de estimulación cognitiva y nos ayudó con la información necesaria para entrenar a nuestra aplicación.

Por otro lado, durante el proceso de pruebas de la aplicación la psicóloga quien hizo uso de la aplicación hizo sugerencias con respecto a la accesibilidad a la aplicación las cuales se mejoraron con la finalidad de que sea accesible.

En cuanto a la fase posterior a las pruebas, al analizar los resultados de las encuestas se pudo observar con gran satisfacción que la aplicación tuvo un porcentaje bastante aceptable en su proceso de generación de subtítulo partiendo de la voz emitida en tiempo real, y que fue de gran apoyo durante las clases de estimulación cognitiva.

6.2. Recomendaciones

Se recomienda enriquecer el corpus de datos para mejorar los modelos, por el momento el sistema solo cuenta con 44 frases entrenadas con dos voces diferentes, por lo que se recomienda expandir el corpus. Al existir un proceso de entrenamiento el sistema se encuentra constantemente aprendiendo, para ello es de gran relevancia extender en gran medida nuestro corpus de datos y buscar que represente los casos más comunes de nuestra lengua, como son acento, entonación, el léxico propio de cada región, con la finalidad que la aplicación pueda ser utilizada en cualquier parte del país.

Al ejecutar el sistema en los sistemas operativos, Linux y Windows podemos recomendar que el sistema sea ejecutado en el sistema operativo Linux debido a que el mismo ofrece mejores tiempos de respuesta.

Para el proceso de actualizar el modelo de lenguaje es necesario que el usuario se conecte a internet, por ello se recomienda que al realizar el proceso de entrenamiento el usuario cargue todos los audios al sistema, a continuación, ejecuta el entrenamiento, el cual se encarga de actualizar el modelo acústico, estos procesos no necesitan de conexión a internet, posteriormente cuando el usuario tenga acceso a internet puede generar el modelo de lenguaje.

6.3. Trabajo futuro

Uno de los trabajos futuros que se ha planteado al trabajar con los adultos mayores fue extender la aplicación y generar los modelos y diccionario necesario para brindar la posibilidad de la identificación futura de idiomas oficiales del Ecuador como son el Quichua y el shuar, apoyando la integración de las diversas comunidades existentes en nuestro país.

También se propone enriquecer el sistema de reconocimiento, por el momento la aplicación se desarrolló sobre una herramienta que trabaja con modelos ocultos de Márkov para la generación de modelos probabilísticos, por lo que a futuro se podría trabajar con redes neuronales para mejorar el proceso de reconocimiento.

Como primera fase se ha propuesto implementarlo en los centros geriátricos con la finalidad de apoyar la accesibilidad del adulto mayor y que el mismo tenga la facilidad de acceder a información y de esta manera mejorar su estilo de vida, este proyecto a futuro puede ser mejorado utilizado en cualquier lugar que brinde servicios a la comunidad pues sería de gran apoyo el poder acceder a las indicaciones no solo escuchándolas, sino que las mismas estarán proyectadas en una pantalla.

REFERENCIAS

- [1] B. B. Villacís and D. J. Carrillo, "Pais Atrevido: La nueva cara sociodemográfica del Ecuador," Edición especial revista Analitika, Quito – Ecuador, 2012.
- [2] C. A. B. M. G. A. A. J. & P. J. S. Cano, "Problemas de la audición en el adulto mayor, factores asociados y calidad de vida: estudio SABE," Bogota, Colombia, 2014.
- [3] M. Cuatlayol and S. F., "Casa de Retiro para el Adulto Mayor," 2012.
- [4] L. Moreno and M. Ribera, "Accesibilidad a los contenidos audiovisuales. Subtitulado," in *XI Congreso Internacional de Interacción Persona-Ordenado*, Valencia, España, 2010.
- [5] O. M. d. I. Salud, "Informe sobre la salud en el mundo," Febrero 2013. [Online]. Available: <https://news.un.org/es/story/2013/02/1265381>. [Accessed Septiembre 2018].
- [6] F. d. e. d. C. d. Andalucía, "Temas para la educación," *Revista Digital para profesionales de la enseñanza*, 2012.
- [7] V. Hernández, F. Castro, R. Belda and J. J. B. de Prat, "Deficiencia, discapacidad y minusvalía auditiva. Auditio," *Revista electrónica de audiolgía*, pp. 19-31, 2006.
- [8] M. d. E. d. Ecuador, "Discapacidades: conceptos elementales," 2018.
- [9] CONADIS, "Consejo Nacional para la Igualdad de Discapacidades," 02 Enero 2018. [Online]. Available: <http://www.consejodiscapacidades.gob.ec/estadistica/index.html>.
- [10] O. m. d. I. Salud, "Envejecimiento y ciclo de vida," [Online]. Available: <http://www.who.int/ageing/about/facts/es/>. [Accessed Septiembre 2018].
- [11] O. M. d. I. Salud, "Discapacidad y rehabilitación," [Online]. Available: http://www.who.int/disabilities/world_report/2011/es/. [Accessed Septiembre 2018].
- [12] E. Villegas, M. Pifarré and E. Santos, "Accesibilidad centrada en el usuario. Creatividad, TICs y sociedad de la información," *Revista Creativa y Sociedad*, 2011.
- [13] S. N. d. I. Discapacidad, "Formatos de Comunicación Accesible Discapacidad Sensorial," 2018.
- [14] Ó. d. J. Á. Chávez, "Guía para crear contenidos digitales accesibles: Documentos, presentaciones, vídeos, audios y páginas web.," *Universidad de Alcalá, Servicio de Publicaciones*, 2015.
- [15] G. Puelles, A. Giancarlo and J. Otoni, "Automatización del proceso de la toma de pedidos en el restaurante DON RULO SAC utilizando una aplicación móvil con reconocimiento de voz soportada por las APPS DE GOOGLE," 2016.
- [16] P. Codosero and J. Manuel, "Desarrollo de un sistema de reconocimiento de habla natural para transcribir contenidos de audio en internet," 2015.
- [17] J. DELLER, J. PROAKIS and J. HANSEN, "Discrete Time Processing of Speech Signals. Discrete Time Processing of Speech Signals Macmillan Publishing," 1993.
- [18] A. Becerra, M. F. Ordóñez and M. Gómez, "Reconocimiento de voz mediante Modelos Ocultos de Markov.," 2009.
- [19] A. Nacelle, "Las redes neuronales: de la biología a los algoritmos de clasificación," 2009.
- [20] C. Luna-Ortega and J. C. Martíñez Romo, "Reconocimiento de Voz con Redes Neuronales, DTW y Modelos Ocultos de Markov," *Conciencia Tecnológica*, 2006.

- [21] J. Celis Nuñez, R. Llanos Castro and Medin, "Modelo Acústico y de Lenguaje del Idioma Español para el dialecto Cucuteño, Orientado al Reconocimiento Automático del Habla.," *Ingeni*, 2017.
- [22] C. Toolkit, 07 Junio 2013. [Online]. Available: https://en.wikipedia.org/wiki/CSLU_Toolkit. [Accessed Septiembre 2018].
- [23] P. Lange and D. Suendermann-Oeft, "Tuning Sphinx to Outperform Google's Speech Recognition API. Conference on Electronic Speech Signal Processing," 2014, pp. 1-10.
- [24] J. Adorf, "Web speech api.KTH Royal Institute of Technology.," 2013.
- [25] CMUSphinx, "Open Source Speech Recognition Toolkit," [Online]. Available: <https://cmusphinx.github.io/>.
- [26] TensorFlow, "Audio Recognition," [Online]. Available: https://www.tensorflow.org/tutorials/sequences/audio_recognition. [Accessed Septiembre 2018].
- [27] P. Schwar, "Phoneme recognizer based on long temporal context," 2019. [Online]. Available: <http://speech.fit.vutbr.cz/software/phoneme-recognizer-based-long-temporal-context>. [Accessed septiembre 2018].
- [28] P. Revuelta, "3. RECURSOS EDUCATIVOS ACCESIBLES EN TIEMPO REAL PARA PERSONAS CON DISCAPACIDAD AUDITIVA SEVERA," *AMADIS 08. III Congreso de accesibilidad a los medios audiovisuales para personas con discapacidad.*, 2009.
- [29] D. Maldonado, R. Villalba and D. Pinto, "Eñe: Sistema de reconocimiento automático del habla en Guaraní," *Simposio Argentino de Inteligencia Artificial (ASAI 2016)-JAIIO 45*, 2016.
- [30] Python, "Python," 2018. [Online]. Available: <https://www.python.org/>.
- [31] Numpu, Fundación Wikimedia, Inc., 20 noviembre 2017. [Online]. Available: <https://es.wikipedia.org/wiki/NumPy>. [Accessed septiembre 2018].
- [32] Matplotlib, Fundación Wikimedia, Inc., 11 noviembre 2017. [Online]. Available: <https://es.wikipedia.org/wiki/Matplotlib>. [Accessed septiembre 2018].
- [33] Q. designer, "Qt Designer Manual," [Online]. Available: <http://doc.qt.io/qt-5/qtdesigner-manual.html>. [Accessed Septiembre 2018].
- [34] D. Boddie, "About PyQt," 06 Junio 214. [Online]. Available: <https://wiki.python.org/moin/PyQt4>. [Accessed Septiembre 2018].
- [35] N. Marco, "Modelado y Programacion 2015-2," [Online]. Available: <https://modelado20152.readthedocs.io/es/latest/pyqt4.html>. [Accessed Septiembre 2018].
- [36] H. Pham, "PyAudio," [Online]. Available: <https://people.csail.mit.edu/hubert/pyaudio/>. [Accessed Septiembre 2018].
- [37] P. S. Foundation, "Struct," 21 Julio 2018. [Online]. Available: <https://docs.python.org/2/library/struct.html>. [Accessed Septiembre 2018].
- [38] P. S. Foundatio, "Wave," 21 Julio 2018. [Online]. Available: <https://docs.python.org/2/library/wave.html>. [Accessed Septiembre 2018].
- [39] M. d. I. E. y. Social, "Norma Técnica Poblacion Adulta Mayor," 28 Febrero 2014. [Online]. Available: <https://www.inclusion.gob.ec/wp-content/uploads/downloads/2014/03/NORMA-TECNICA-DE-POBLACION%CC%81N->

ADULTA-MAYOR-28-FEB-2014.pdf. [Accessed 2017].

ANEXOS

- **Modelo de encuesta aplicada**

Encuesta sobre la utilización de la aplicación de escritorio “Capturador de audio para extracción de subtítulo en línea como herramienta de apoyo para personas de tercera edad y aquellas que presentan discapacidad auditiva” a los residentes del “Centro Gerontológico Rosa Elvira de León”

La siguiente encuesta será llenada en colaboración la con el apoyo de la Psicóloga del centro geriátrico para responder las preguntas que hemos elaborado con la intención de saber cuál es la percepción del adulto mayor respecto a la utilidad y aplicación de un sistema que captura el habla extrae el subtítulo y se presenta en pantalla, para apoyar el proceso de estimulación cognitiva.

Nombres y Apellidos: _____

Edad en años cumplidos: _____

Nivel de Instrucción: _____

Estado civil:

Soltero/a

Viudo/a

Casado/a

Unión libre

Divorciado/a

Otra (indique por favor): _____

Indique en qué trabajó: _____

Indique hace cuánto tiempo se jubiló: _____

Indique por qué se jubiló: _____

1. ¿Tiene discapacidad auditiva?

Sí

No

2. Condición auditiva que presenta:

Sordera Unilateral (Pérdida auditiva en un solo oído)

Sordera Bilateral (Pérdida auditiva en ambos oídos)

Pérdida auditiva post-lingüística (Sucedió después de aprender a hablar)

Pérdida auditiva prelingüística (Sucedió antes aprender a hablar)

Sordera Gradual (La pérdida auditiva empeoró con el paso del tiempo)

Sordera Repentina (La pérdida auditiva ocurrió repentinamente)

Sordera Congénita (La pérdida auditiva estuvo presente desde el momento del nacimiento)

Otra(especifique): _____

ENTORNO AUDIOVISUAL DE LA APLICACIÓN

3. El tamaño de los botones que posee la aplicación es:

Absolutamente grande

Muy grande

Suficiente

Pequeño

Muy pequeño

4. En la siguiente tabla indique: La percepción sobre el tamaño de letra aplicado al subtítulo que se presenta en la pantalla.

	Valoración Para el tamaño de Fuente San-Serif
--	--

TAMAÑO	Totalmente legible	Muy legible	Legible	Poco legible	Nada legible
60 pts					
75 pts					
90 pts					

5. En la siguiente tabla indique: La percepción sobre la combinación de colores del entorno para el subtitulado.

Valoración Para la combinación de Colores fondo y Fuente					
FONDO/ FUENTE	Absolutamente de acuerdo	Muy de Acuerdo	De acuerdo	Poco de acuerdo	Absolutamente en desacuerdo
Blanco/ Negro					
Negro/Blanco					
Negro/Amarillo					
Amarrillo/Negro					
Azul/Blanco					
Celeste/Blanco					

Entorno Funcional de la Aplicación

6. En la siguiente tabla indique: La percepción respecto a la comprensión de las frases que el sistema presenta en pantalla.

Categoría:	Orientación Tiempo, Espacio, Persona.				
FRASES:	Totalmente entendibles	Muy entendibles	Entendibles	Poco entendibles	Nada entendibles
¿Qué día es hoy?					
¿En qué fecha estamos?					
¿Cómo se llama este centro?					
¿Como se llama la directora de este centro					
¿Como se llama la calle					
¿Qué queda al frente					
¿Cómo se llama la ciudad?					
¿Cómo se llama la provincia?					
¿Cómo se llama el país?					
Categoría:	Lenguaje				
FRASES	Totalmente entendibles	Muy entendibles	Entendibles	Poco entendibles	Nada entendibles

Encantado de conocerle					
Es un placer saludarle					
¿Cómo está usted?					
Hasta luego					
Hasta Pronto					
Hasta mañana					
Me puede decir dónde queda esta calle					
Estoy muy agradecido					
Tengo Hambre					
Tengo Calor					
Categoría:	Memoria, Atención				
FRASES	Totalmente entendibles	Muy entendibles	Entendibles	Poco entendibles	Nada entendibles
El que madruga Dios le Ayuda					
No hay mal que por bien no venga					
Mas vale estar solo que mal acompañado					
Caer está permitido levantarse es obligatorio					
A buena hambre no hay pan duro					
A buen entendedor pocas palabras bastan					
Categoría:	Funciones ejecutivas				
FRASES	Totalmente entendibles	Muy entendibles	Entendibles	Poco entendibles	Nada entendibles
Ya está listo el refrigerio					
Ya le tengo que bañar					
Ya le tengo que cepillar los dientes					
Ya le tengo que cortar las uñas					
Ya le tengo que lavar las manos					
Venga a tomar café					

Vamos a misa					
Subamos a merendar					
Es hora del desayuno					
Es hora de visitar al doctor					

7. ¿Qué opina sobre el tamaño que tiene en la pantalla?:

- Demasiado grande
- Grande
- Suficiente
- Un poco pequeña
- Demasiado pequeña

8. ¿Qué opina sobre la aplicación para adultos mayores?:

- Absolutamente indispensable
- Muy necesario
- Necesario
- Poco necesario
- No es necesario

9. ¿Qué opina sobre usar esta aplicación para sus actividades diarias?:

- Me ayuda muchísimo para comunicarme
- Me ayuda mucho para comunicarme
- Me ayuda a comunicarme
- No me ayuda mucho a comunicarme
- No me ayuda en nada para comunicarme

10. ¿Qué piensa usted sobre las frases que observó en la pantalla?:

- Son totalmente interesantes
- Son muy interesantes
- Son interesantes
- No me interesan mucho
- No me interesan en lo absoluto

11. ¿Qué le parecen las frases que observó en la pantalla?:

- Son absolutamente fáciles de comprender
- Son muy fáciles de comprender
- Son fáciles de comprender
- Son un poco difíciles de comprender
- No comprendo nada

12. ¿Qué piensa sobre la aplicación para alfabetizar a los adultos mayores?

- Es absolutamente útil
- Es muy útil
- Es útil
- Es poco útil

No es nada útil

13. ¿Después las terapias con la aplicación como se siente?:

- Comprendo totalmente las cosas
- Comprendo bastante bien las cosas
- Comprendo las cosas
- No comprendo casi nada
- No comprendo nada

14. ¿Qué opina sobre la utilidad de aplicación para la comunicación de los adultos mayores?:

- Es excelente
- Es muy buena
- Es buena
- No es muy buena
- Es pésima

- **Certificado de Participación en el IX Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2018) - VI Conferencia Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones para mejorar la Accesibilidad (ATICAcces 2018).**





- **Consentimiento informado, firmado por la Psicóloga del Centro gerontológico “Rosa Elvira de León”**



CONSENTIMIENTO INFORMADO PARA ADULTOS MAYORES Y PERSONAS CON DISCAPACIDAD AUDITIVA.

UNIVERSIDAD POLITÉCNICA SALESIANA
Facultad de Ingeniería de Sistemas

Grupo de Investigación en Inteligencia Artificial y Tecnologías de Asistencia
Cátedra UNESCO

CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD AUDITIVA

CONSENTIMIENTO INFORMADO

INSTRUCCIONES: Este consentimiento informado luego de ser leído por las personas que participarán en el proyecto será firmado.

Usted está invitada(o) a participar en un programa de investigación diseñado para obtener información científica que puede ayudar a otras instituciones en el futuro.

El objetivo de esta investigación es determinar la percepción sobre la funcionalidad de la aplicación “CAPTURADOR DE AUDIO PARA EXTRACCIÓN DE SUBTITULADO EN LÍNEA COMO HERRAMIENTA DE APOYO PARA PERSONAS DE TERCERA EDAD Y AQUELLAS QUE PRESENTAN DISCAPACIDAD AUDITIVA” como apoyo para el acceso a la información.

En el estudio participarán las mujeres y hombres de tercera edad entre 60 a 95 años que presenten o no discapacidad auditiva residentes del Centro gerontológico “Rosa Elvira de León” en la Ciudad de Azogues bajo la supervisión de los autores de esta investigación.

La decisión de participar de cada una de las personas es voluntaria. Si por alguna razón usted decide no participar de este estudio, esta decisión no afectará la relación con los investigadores y autoridades institucionales.

La participación en el estudio implica dar las facilidades para que los investigadores tengan acceso a la información producida mediante las encuestas. Este estudio no representa riesgo de ningún tipo para las participantes. La información que usted nos facilite es absolutamente confidencial.

No incluiremos ninguna información que pueda hacer posible la identificación personal en publicaciones o reporte posteriores, será todo anónimo.

No hay costos para usted ni para la institución por tomar parte en esta investigación, ni tampoco se entregará ningún incentivo al personal por su participación.

Por favor, tómese su tiempo para decidir y pedir explicaciones sobre cualquier palabra o información que no entienda a los directivos del proyecto. Para obtener mayor información sobre el estudio puede dirigirse a los responsables de la investigación Ing. Paola Cristina Ingavélez Guerra o a la estudiante de Ingeniería de Sistemas Inés Patricia Yambay Aulla.

(Contacto Ing. Paola Cristina Ingavélez Guerra: 0987958807, Inés Patricia Yambay Aulla: 0992633576).

Usted recibirá una copia de este consentimiento.

Declaración de Consentimiento:

He leído atentamente y entiendo que puedo hacer preguntas sobre el estudio y estoy de acuerdo con las respuestas. Voluntariamente acepto participar en este estudio y entiendo que cualquier persona tiene el derecho de retirarse en cualquier momento sin que esto signifique ningún perjuicio para mí.

Firmado este consentimiento no delego ningún derecho legal que me pertenezca.

Saskia Andrade Ordóñez

Nombre

Saskia Andrade

Firma



A circular stamp is located to the right of the signature. The text within the stamp, arranged in a circular pattern, reads: "CENTRO GERONTOLÓGICO DE LA UNIVERSIDAD DE LEÓN, FACULTAD DE CIENCIAS SOCIALES".