Prototipo de aplicación móvil en Android para transcripción de archivos de audio a texto en idioma español dirigida a personas con discapacidad auditiva

German Leonardo Rey Benitez - Jhon Alexander Morales Gaitan - Luis Abelardo Melo Melo

Programa Ingeniería de Sistemas, Fundación Universitaria Compensar

Proyecto fin de grado para Facultad de Ingeniería Sistemas

José Luis Cabra Lopez

Noviembre 13 del 2023

Esta obra está bajo una Licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional.

Resumen

Teniendo presente la poca presencia de aplicaciones de transcripción en idioma español en los portales oficiales de distribución de aplicaciones móviles para Android como Google Play Store y Aurora Store presentamos la construcción e implementación de un prototipo de aplicación móvil en Android el cual integra una de las más reconocidas herramientas de inteligencia artificial y red neuronal llamada ***Whisper*** del pionero ***OPENAI***, el cual se acerca a la robustez y precisión a nivel humano en el reconocimiento de voz ofreciendo un 97% de exactitud en la transcripción permitiendo la extracción de texto en español desde un archivo de audio con el objetivo de proporcionar un servicio a las personas con discapacidad auditiva para el consumo de audio digital.

Palabras clave: Aplicación, Android, Extracción, Sordera, Transcripción., Whisper, OpenAI

Abstract

Taking into account the limited presence of transcription applications in Spanish in the official portals for the distribution of mobile applications for Android such as Google Play Store and Aurora Store, we present the construction and implementation of a mobile application prototype on Android which integrates one of the most recognized artificial intelligence and neural network tools called Whisper from the pioneer OPENAI, which approaches human-level robustness and precision in voice recognition, offering 95% accuracy in transcription, allowing the extraction of text in Spanish from a file audio with the aim of providing a service to people with hearing disabilities to the consumption of digital audio.

Keywords: *Application, Android, Extraction, Deafness,* Transcription, Whisper, OpenAI

Introducción

En el presente proyecto, se aborda la creación de un prototipo de aplicación móvil en Android con el objetivo de brindar apoyo a las personas con discapacidad auditiva. Este prototipo de aplicación que llamaremos ***TranscripApp*** se enfoca específicamente en la extracción de texto en español a partir de archivos de audio digital, buscando facilitar la comunicación y el acceso a la información para este grupo de personas que requieran transcribir audio a texto.

La importancia de este proyecto radica en la necesidad de motivar la inclusión y la igualdad de oportunidades para las personas con discapacidad auditiva. La falta de accesibilidad a la información puede limitar su participación en la sociedad y dificultar su comunicación con personas que no conocen el lenguaje de señas. Nuestra investigación previa al desarrollo TranscripApp evidencio que este tipo de aplicaciones inclusivas no tiene una fuerte presencia en el mercado actual en el idioma español, al desarrollar un prototipo de aplicación móvil que permita la extracción de texto a partir de archivos de audio, se busca proporcionar una herramienta que permita a las personas con discapacidad auditiva comprender y participar de manera más efectiva en diferentes situaciones y contextos. Los antecedentes teóricos y prácticos de este proyecto se basan en el avance de la tecnología de reconocimiento de voz (ASR) y la capacidad de las aplicaciones móviles para procesar y convertir el habla garbado en medios digitales en texto. Existen investigaciones previas que han demostrado la viabilidad y efectividad de estas tecnologías en otros idiomas, pero es necesario adaptarlas y evaluarlas para el español.

Los objetivos de este proyecto incluyen la implementación de un algoritmo de reconocimiento de voz en español tomando como insumo la red neuronal ***Whisper*** de OPENAI, la integración de este algoritmo en una app móvil para dispositivos Android permitiendo extracción de texto a partir de archivos de audio desde grabaciones de voz en diferentes entornos y condiciones. La aplicación móvil integrara un diseño simple fácil de usar permitiendo a los usuarios cargar archivos de audio y obtener el texto correspondiente de manera rápida y precisa.

Sin embargo, es importante mencionar las limitaciones de este proyecto. La precisión de la extracción de texto dependerá de la calidad del audio y la claridad del habla. En situaciones con ruido de fondo o habla poco clara, la exactitud puede verse afectada. Además, la aplicación se limitará a la extracción de texto en español y no contemplará otros idiomas. El diseño de interfaces de usuario amigables, la implementación de la funcionalidad de extracción de texto y la realización de pruebas y evaluaciones para medir su precisión y desempeño.

El significado que este estudio tiene en el avance del campo respectivo radica en la posibilidad de mejorar la calidad de vida y la inclusión social de las personas con discapacidad auditiva facilitando su participación en diferentes ámbitos, como la educación, la información y la comunicación, fortaleciendo su autonomía y empoderamiento.

En cuanto a su aplicación nuestro prototipo de aplicación móvil TranscripApp busca cubrir una brecha existente en la accesibilidad para personas con discapacidad auditiva. El desarrollo de esta herramienta tecnológica puede abrir nuevas oportunidades e impulsar investigaciones futuras en el campo de la accesibilidad para personas con discapacidad, tanto a nivel tecnológico como social en el habla hispana.

Planteamiento y justificación

La Organización Mundial de la Salud (OMS) utiliza los términos "sordera" y "pérdida de audición" para referirse a la disminución de la capacidad auditiva de una persona, en su artículo "*Sordera y pérdida de la audición" de 28 Jul, 2023*, define la pérdida de audición como la incapacidad de oír tan bien como una persona con un umbral de audición normal, que es igual o mejor que 20 dB en ambos oídos. La pérdida de audición puede ser leve, moderada, grave o profunda, y puede afectar a uno o ambos oídos. Las personas "duras de oído" son aquellas cuya pérdida de audición es entre leve y grave, mientras que las personas "sordas" padecen una pérdida de audición profunda, lo que significa que oyen muy poco o nada. La OMS

estima que más del 5% de la población mundial, es decir, 360 millones de personas, tienen pérdida de audición discapacitante (p, *Organización Mundial de la Salud, 2023*) y uno de los desafíos más grandes que tienen las personas con discapacidad auditiva o reducción de audición es comprender y consumir archivos multimedia como lo son audios enviados a través de aplicaciones o plataformas móviles de mensajería instantánea, lo cual plantea limitaciones y barreras para el uso de la tecnología, recordemos que esta sociedad es caracterizada por el uso de tecnología llamándose como "Era Digital" y el problema que afrontan millones de personas es algo inverosímil que merece todo el apoyo y enfoque en búsqueda de soluciones o alivio; con el desarrollo de nuestro prototipo de aplicación móvil las personas con discapacidad auditiva o sordera podrán tener acceso a un desarrollo y uso complementario de la tecnología a la mano como son los smartphones.

Objetivo General

Implementar un prototipo funcional de una aplicación móvil en Android que permita la transcripción automática de audio a texto en español incorporando herramientas de aprendizaje automático e inteligencia artificial.

Objetivos Específicos - Planeación

1. **Implementar el prototipo de la aplicación de transcripción:**
    - Implementar una interfaz gráfica de usuario al prototipo funcional de la aplicación móvil en Android.
2. **Evaluación y mejora desarrollo técnico de la herramienta de transcripción:**
    - Analizar y evaluar la precisión de la herramienta de transcripción a través de pruebas teniendo en cuenta indicadores de cumplimiento para la funcionalidad principal de transcripción de audio a texto.
3. **Evaluación de funcionamiento:**
        - Realizar pruebas finales de funcionamiento para evaluar la funcionalidad de la aplicación en términos de navegación, comprensión y usabilidad general.

Antecedentes y Justificación

En los últimos años, se han llevado a cabo diversos estudios e investigaciones relacionados con el reconocimiento automático de voz, la transcripción de texto y la síntesis de voz. Estos avances tecnológicos han permitido desarrollar sistemas y herramientas que facilitan la comunicación y el acceso a la información para diferentes grupos de usuarios.

Por ejemplo, se ha investigado sobre el análisis de texto y el procesamiento de sistemas de conversión de texto a voz en diferentes idiomas, como el tailandés *(Xuee Lin et al., 2014)*. También se han realizado estudios sobre la mejora de la legibilidad en la transcripción con reconocimiento automático de voz *(Jun Wei Liao, 2023)* y sobre el reconocimiento de voz a texto y texto a voz mediante aprendizaje profundo *(Rumia Sultana, 2014)*.

Además, se han explorado diferentes enfoques para la traducción de voz, incluyendo la traducción en cascada o directa *(Thierry Etchegoyhen, 2022)* y el desarrollo de corpus de traducción de voz a texto *(Maria Labied, 2023)*. También se ha investigado sobre el reconocimiento automático de voz en idiomas de bajos recursos es decir con recursos de datos limitados en sus vocablos y poca expresión del lenguaje *(Thomas Reitmaier, 2022)* y técnicas de mejora del habla basadas en el procesamiento de señales de audio *(Caleb Rascón, 2023).*

Estos estudios y avances tecnológicos han sentado un precedente para el desarrollo de nuestro proyecto. Teniendo en cuenta la necesidad de mejorar la comunicación y el acceso a la información para las personas con discapacidad auditiva, hemos decidido enfocarnos en el desarrollo de un prototipo de aplicación móvil que permita la extracción de texto en español a partir de archivos de audio usando ***TranscripApp***. Porque se ha evidenciado que existen herramientas de transcripción actualmente como *Talk to deaf de Kickdata*, o *Ava: Transcriptions & Captions de Ava Accessibility*, pero estas aplicaciones o similares son de bajo desarrollo para la lengua española esto se puede evidenciar en su perfil en la Google Play Store, además su

desarrollo esta relegado u olvidado en relación con el habla hispana teniendo un mayor enfoque en otros idiomas.

Además, este proyecto se basa en los avances tecnológicos como la red neural *Whisper* de OPENAI y las investigaciones previas en el campo del reconocimiento automático de voz (ASR) y el procesamiento del habla. Se ha demostrado la viabilidad y efectividad de estas tecnologías en otros idiomas como el Ingles Según *International Speech Communication Association-ISCA* en su Newsroom (p, *www.isca-speech.org*) pero nuestro proyecto apunta a su adaptación y despliegue para el idioma español.

Marco Histórico

El uso de la tecnología para ayudar a las personas con discapacidad auditiva ha evolucionado a lo largo del tiempo. A continuación, se presentan algunos hitos importantes en la historia del uso de la tecnología para personas con discapacidad auditiva:

- En la década de 1960, se desarrollaron los primeros audífonos electrónicos, que mejoraron significativamente la calidad de vida de las personas con discapacidad auditiva.
- En la década de 1980, se crearon los primeros sistemas de bucle magnético, que permitían a las personas con discapacidad auditiva escuchar mejor en lugares públicos, como teatros y salas de conferencias.
- En la década de 1990, se desarrollaron los primeros sistemas de subtitulado en tiempo real para televisión, que permitían a las personas con discapacidad auditiva seguir los programas de televisión en tiempo real.
- En la década de 2000, se crearon los primeros sistemas de interpretación de lengua de signos en videollamadas, que permitían a las personas con discapacidad auditiva comunicarse con otras personas a través de la lengua de signos.

- En la actualidad, existen numerosas tecnologías que ayudan a las personas con discapacidad auditiva, como las aplicaciones móviles que ofrecen información sobre espacios culturales accesibles para personas sordas y con discapacidad auditiva, los sistemas de detección de sonidos y eventos sonoros del hogar, y los guantes de comunicación de lenguaje de gestos (*BrightSign, Hadeel Ayoub*).

En general, la tecnología ha sido un gran aliado para las personas con discapacidad auditiva, ya que les ha permitido mejorar su calidad de vida permitiendo su desarrollo laboral, profesional e inclusión social.

Marco Conceptual:

El proyecto para la concepción y desarrollo del prototipo de aplicación móvil ***TranscripApp*** se basa en la accesibilidad, la inclusión y el desarrollo de tecnologías que ayuden a las personas con discapacidad auditiva a superar las barreras de comunicación y a tener una participación plena en la sociedad. Algunos aspectos clave a considerar son:

- **Accesibilidad:** La accesibilidad se refiere a la capacidad de las personas con discapacidad para acceder y utilizar la tecnología de manera efectiva. Esto implica diseñar soluciones que sean comprensibles, utilizables y navegables para todos, incluidas las personas con discapacidad auditiva.
- **Inclusión:** La inclusión implica garantizar que las personas con discapacidad tengan igualdad de oportunidades y puedan participar plenamente en la sociedad. El desarrollo de tecnologías accesibles es fundamental para lograr la inclusión de las personas con discapacidad auditiva.
- **Derechos humanos:** Las personas con discapacidad tienen derechos humanos fundamentales, incluido el derecho a la igualdad de oportunidades, la no discriminación y la participación en la vida cultural, social y política. El marco jurídico internacional reconoce y protege estos derechos.

- **Tecnología de asistencia:** La tecnología de asistencia es un conjunto de dispositivos o equipos que ayudan a las personas con discapacidad a participar plenamente en las actividades de la vida diaria. En el caso de las personas con discapacidad auditiva, la tecnología de asistencia puede incluir dispositivos de transcripción de audio a texto en español, servicios de interpretación en lengua de signos y sistemas de alerta visual.

- **Desarrollo sostenible:** El desarrollo de tecnologías accesibles y la promoción de la inclusión de las personas con discapacidad son aspectos clave del desarrollo sostenible. Estos proyectos contribuyen a romper las barreras digitales y a garantizar que nadie se quede atrás en la era de la tecnología.

Marco Legal

En Colombia, existen diversas normativas y regulaciones que buscan garantizar la accesibilidad y la inclusión de las personas con discapacidad auditiva en el uso de la tecnología. Estas normativas y regulaciones se basan en los derechos humanos y la accesibilidad. El proyecto del prototipo de aplicación móvil TranscripApp para la transcripción de audio a texto español y la ayuda a personas con discapacidad auditiva debe cumplir con las normativas y regulaciones existentes en Colombia. Algunas de las normativas y regulaciones que se deben considerar son:

- Ley 982 de 2005, esta ley establece el derecho de las personas sordas o sordociegas a acceder a una forma de comunicación, ya sea la Lengua de Señas Colombiana o el oralismo. Además, la ley establece que los teléfonos públicos deben contar con características técnicas que permitan a los limitados sensoriales el acceso a este servicio.

- El Instituto Nacional para Sordos (INSOR), establece y promociona en su página web (p, www.insor.gov.co) la "Guía para la producción de contenidos audiovisuales con ajustes razonables para la población sorda colombiana" como publicación oficial la cual es actualizada anualmente.

- Convención Suramericana de Personas Sordas, esta convención, organizada por la Secretaría Regional de la Federación Mundial de Sordos y la Federación Nacional de Sordos de Colombia (FENASCOL), ha sido un espacio para la promoción de la inclusión y la igualdad de oportunidades para las personas sordas en Colombia.

- Ley 1341 de 2009, esta ley establece el marco legal para el desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) en Colombia. En este marco legal se establecen los principios de accesibilidad y usabilidad de las TIC para garantizar la inclusión de todas las personas, incluidas las personas con discapacidad auditiva.

Marco Teórico

La tecnología digital se ha convertido en una herramienta importante para la vida diaria de las personas con discapacidad auditiva. Las tecnologías digitales pueden ayudar a las personas con discapacidad auditiva a acceder a la información, participar en actividades sociales y a desenvolverse en sus vidas de manera más efectiva.

La tecnología también puede ayudar a las personas con discapacidad auditiva a superar las barreras de comunicación. Por ejemplo, los sistemas de subtitulado en tiempo real para televisión y las aplicaciones móviles que ofrecen información sobre espacios culturales accesibles para personas sordas y con discapacidad auditiva son herramientas útiles para superar las barreras de comunicación, el uso de las TIC en la educación de las personas con discapacidad auditiva puede mejorar su rendimiento académico y su

inclusión social; Las TIC pueden ayudar a los estudiantes con discapacidad auditiva a acceder a la información y a participar en el aula de manera más efectiva.

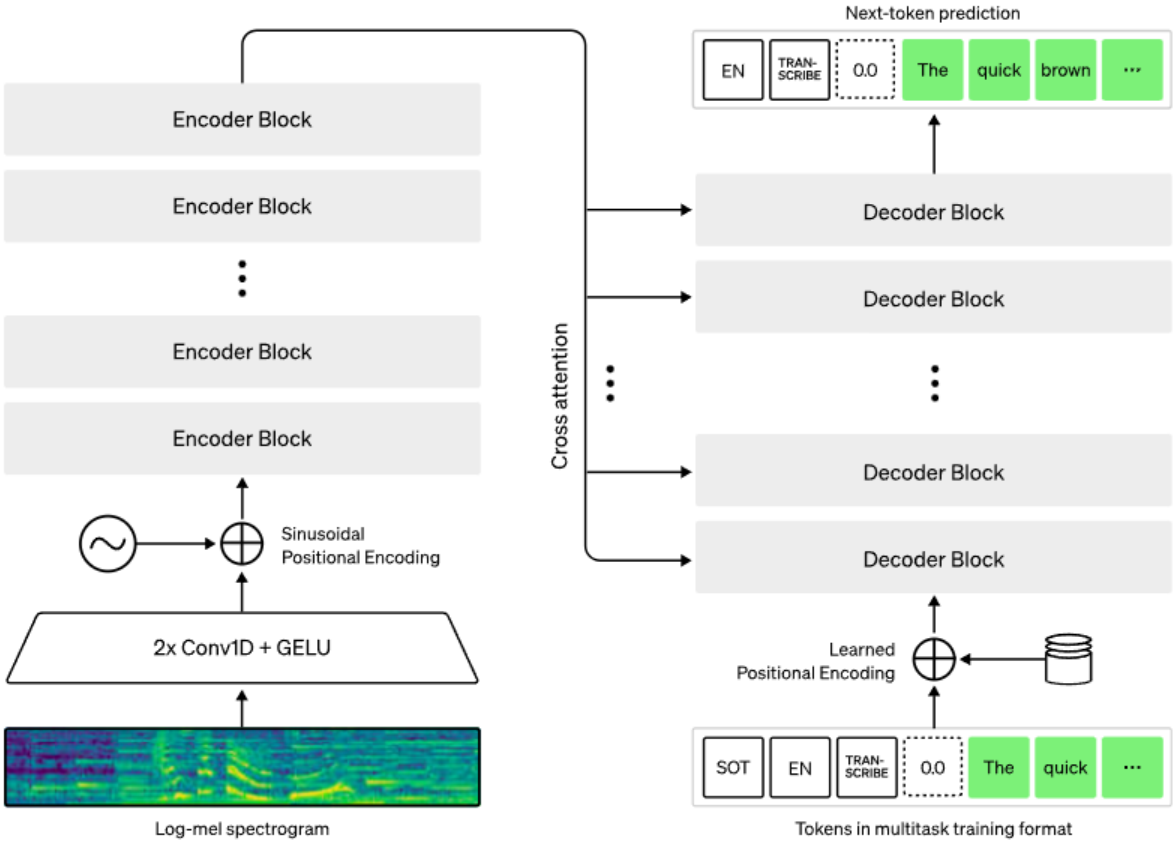
La población objetivo de este proyecto son las personas con discapacidad auditiva, que enfrentan barreras de comunicación y acceso a la información en su vida diaria. Estas personas pueden tener dificultades para comunicarse con otras personas, acceder a servicios y recursos, y participar plenamente en la sociedad. La aplicación móvil para la transcripción de audio a texto español y la ayuda a personas con discapacidad auditiva busca mejorar la accesibilidad y la inclusión de estas personas, permitiéndoles acceder a la información y comunicarse de manera efectiva.

Esta población incluye a personas de todas las edades y géneros que presentan discapacidad auditiva, ya sea de nacimiento o adquirida a lo largo de su vida. Estas personas pueden tener diferentes niveles de discapacidad auditiva, desde una pérdida leve hasta una pérdida total de la audición. Además, la población objetivo puede incluir a personas que utilizan la lengua de señas como su principal forma de comunicación, así como a personas que utilizan otros métodos de comunicación, como la lectura de labios o la escritura.

Nuestro prototipo de aplicación móvil ***TranscripApp*** la cual integra tecnología **ASR** - En inglés conocido como "*Automatic Speech Recognition*", que es una tecnología de asistencia que consiste la detección, captura, extracción, modelado acústico, decodificación y procesamiento de fonemas y patrones en el habla para la transcripción de audio a texto en español y la ayuda a personas con discapacidad auditiva puede ser útil para personas en diferentes situaciones, en el reconocimiento automático de la voz como en el trabajo, en la escuela, en el hogar o en situaciones sociales. La aplicación podría ser utilizada por personas que necesitan transcripciones de conversaciones, reuniones, conferencias, clases, entre otros. Además, la aplicación puede ser utilizada para acceder a información en línea, como noticias, artículos, podcast, grabaciones, entre otros los formatos compatibles de audio digital son: mp3, mp4, mpeg, mpga, m4a, wav, y webm.

La implementación de este proyecto incluye diferentes tareas, desde la implementación y adecuación de herramientas ya existentes y el estado del arte, hasta el diseño y desarrollo de la interfaz de usuario para el procesamiento del audio. En la siguiente figura podemos ver el esquema del funcionamiento básico de la aplicación a través del uso del algoritmo y red neuronal *Whisper* de OpenAI.

Figura 1 *Esquema de funcionamiento básico del algoritmo Whisper OpenAI*



***Nota*:** Tomado de OpenAI -Whisper, https://openai.com/research/whisper

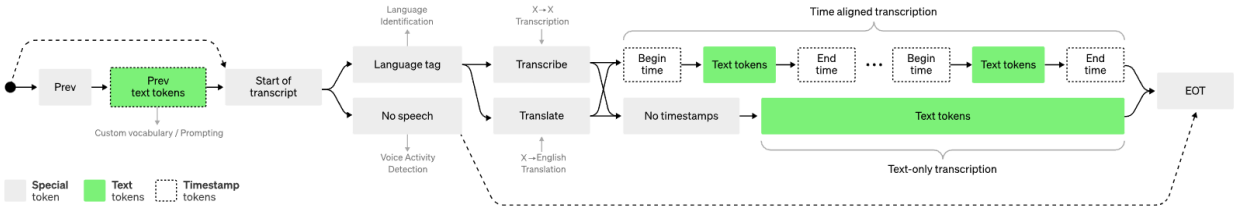
Según el portal oficial de OPENAI "***Whisper*** es un sistema de reconocimiento automático de voz (ASR-Automatic Speech Recognition) entrenado por su casa matriz OPENAI con más de 680.000 horas de datos examinados y multifinalidad recopilados de la web. Demostrando que el uso de un conjunto de datos muy grande y diverso conduce a una fuerte fuente a los acentos, el ruido de fondo y el lenguaje técnico. Además,

permite la en varios idiomas, así como la traducción de esos idiomas al inglés. El código de los modelos y el código de inferencia sirven de base para la creación de aplicaciones útiles y para futuras investigaciones sobre el procesamiento robusto del habla.”

(p, OPENAI, https://openai.com/research/whisper)

“La arquitectura de *Whisper* es un enfoque simple de extremo a extremo, implementado como un transformador codificador-decodificador. El audio de entrada se divide en fragmentos de 30 segundos, se convierte en un espectrograma *log-Mel* y luego se pasa a un codificador. Un decodificador está entrenado para predecir el título de texto correspondiente, entremezclado con tokens especiales que dirigen al modelo único para realizar tareas como la identificación de idiomas, las marcas de tiempo a nivel de frase, la transcripción de voz al español.” (p, OPENAI, https://openai.com/research/whisper)

Figura 2 *Diagrama de operación del Red Neuronal Whisper de OpenAI*



***Nota*:** Tomado de OpenAI -Whisper, https://openai.com/research/whisper

Estado del Arte

Tema	Antecedentes	Tendencias	Líneas de investigación	Fuentes
Discapacidad auditiva	La discapacidad auditiva es una condición que afecta a la capacidad de una persona para escuchar. Puede ser causada por una variedad de factores, incluyendo nacimiento, enfermedad o lesión. Las personas con discapacidad auditiva pueden experimentar una variedad de dificultades, incluyendo dificultades para comunicarse, aprender y participar en la sociedad.	El número de personas con discapacidad auditiva en el mundo está aumentando. En Colombia, se estima que el 4,5% de la población tiene algún grado de discapacidad auditiva.	Investigación en el campo de la rehabilitación de personas con discapacidad auditiva. Investigación en el campo de la tecnología para personas con discapacidad auditiva.	[7] [1] [2]
Transcripción de audio a texto	La transcripción de audio a texto es un proceso que convierte el habla en texto escrito. Esta tecnología puede ser utilizada para ayudar a las personas con discapacidad auditiva a acceder a la información que se encuentra en audio.	Los sistemas de reconocimiento automático del habla (ASR) han mejorado significativamente su precisión en los últimos años.	Investigación en nuevos algoritmos de ASR. Investigación en el uso de ASR en aplicaciones móviles.	[3] [4] [7] [8] [15] [16]
Reconocimiento automático del habla (ASR)	El reconocimiento automático del habla es una rama de la inteligencia artificial que permite a los sistemas informáticos reconocer el habla humana.	Los sistemas de ASR se utilizan en una variedad de aplicaciones, incluyendo la transcripción de audio a texto, la traducción automática y el control por voz.	Investigación en nuevos algoritmos de ASR que sean más precisos y robustos. Investigación en el uso de ASR en aplicaciones de nueva generación.	
Accesibilidad	La accesibilidad es el grado en que las personas con discapacidad pueden acceder y utilizar un producto o servicio.	La accesibilidad es un tema importante de interés para los gobiernos, las empresas y las organizaciones de la sociedad civil.	Investigación en nuevos estándares y directrices de accesibilidad. Investigación en el uso de tecnologías accesibles para personas con discapacidad.	[8] [9] [10]

Metodología propuesta

Teniendo como base fundamental el objetivo general del proyecto planteado se propone el plan de acción y metodología así:

1. Identificar los objetivos del proyecto.
2. Definir los requisitos del proyecto.
3. Seleccionar las tecnologías y herramientas adecuadas para el proyecto.
4. Diseñar la arquitectura del proyecto.
5. Detallar cada paso el desarrollo del proyecto.
6. Realizar pruebas y correcciones documentadas.
7. Lanzar el proyecto (Documentación y Aplicación).

Cronograma

El marco de trabajo utilizado fue SCRUM el cual se basa en un enfoque iterativo e incremental, aquí un detalle de los sprints utilizados:

Tabla 1 *Sprints proyecto prototipo de aplicación móvil TranscripApp.*

Sprints- Actividades	Semana													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sprint 1:														
Definir los requisitos y objetivos del proyecto.	X													
Realizar una investigación de mercado y análisis de competencia.		X												
Diseñar la arquitectura de la aplicación y establecer las tecnologías a utilizar.			X											
Crear los primeros prototipos de la interfaz de usuario.				X										
Sprint 2:														
Desarrollar el flujo de registro y inicio de sesión del prototipo de la aplicación.					X									
Integrar librerías o servicios externos necesarios para el proyecto.						X								
Realizar pruebas unitarias y corregir posibles errores.							X							
Sprint 3:														
Implementar la funcionalidad principal de la aplicación.								X	X					
Realizar pruebas funcionales y de integración.										X	X			
Optimizar el rendimiento y la usabilidad de la aplicación.												X		
Sprint 4:														
Mejorar y expandir la funcionalidad del prototipo de la aplicación según las necesidades identificadas.													X	
Realizar pruebas de rendimiento y realizar ajustes necesarios.													X	
Añadir características adicionales según los comentarios y retroalimentación de los usuarios.														X
Realizar pruebas exhaustivas del prototipo de la aplicación antes de su lanzamiento.														X

***Nota*:** Elaboración propia.

Evaluación

La evaluación del proyecto se llevará a cabo en una fase:

Evaluación interna: La evaluación interna se llevará a cabo por el equipo de desarrollo de software. Se utilizarán los siguientes indicadores: Cantidad de palabras incorrectas, lapso, Tiempo de respuesta en Milisegundos, Porcentaje de Error, Porcentaje de exactitud los cuales nos sirven para evaluar la precisión y la usabilidad del prototipo. Los resultados de la evaluación se utilizarán para mejorar el prototipo y para determinar si se puede llevar a cabo una implementación a gran escala.

Consideración especial para el desarrollo en Colombia

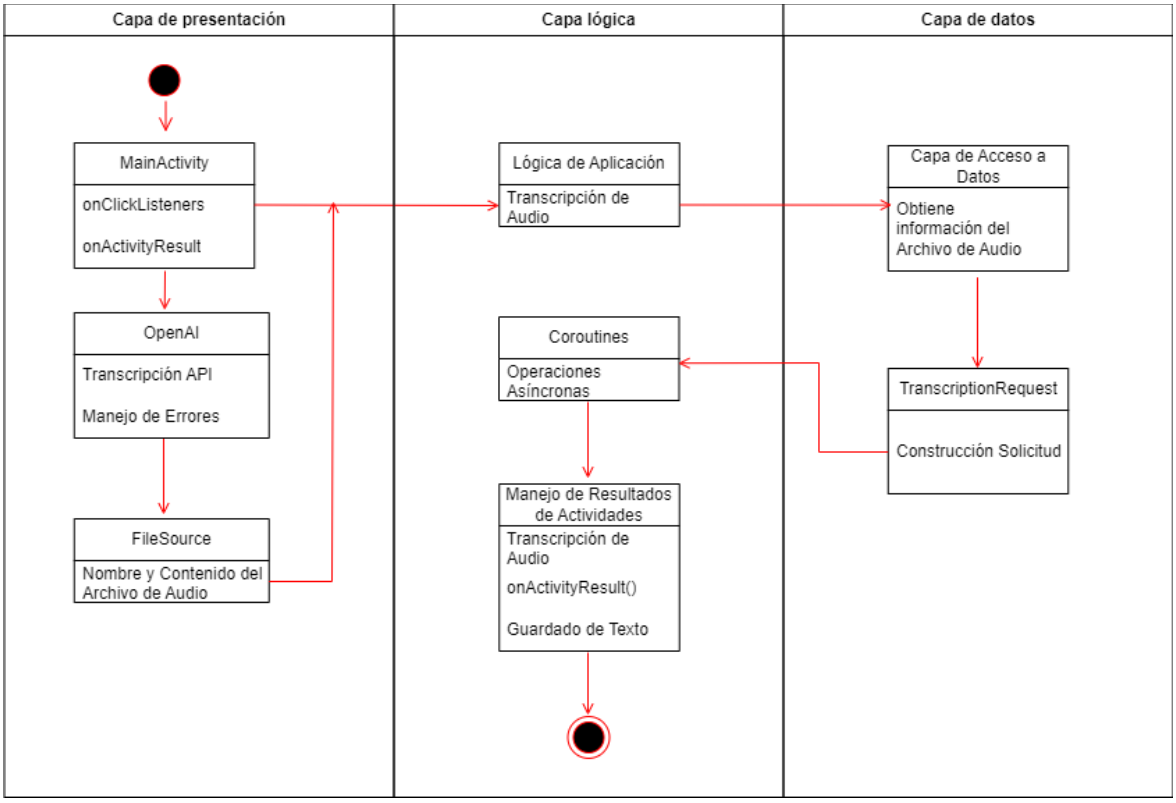
El proyecto se desarrollará en Colombia, por lo que se deben tener en cuenta las siguientes consideraciones especiales:

- **La diversidad lingüística:** El proyecto se centrará en la transcripción de audio a texto en idioma español. Sin embargo, es importante tener en cuenta la diversidad lingüística de Colombia, que incluye una variedad de dialectos y lenguas indígenas.

El equipo de desarrollo del proyecto trabajará para superar estos desafíos y garantizar que el prototipo sea accesible y útil para las personas con discapacidad auditiva en Colombia.

Diagrama de Software

Figura 3 *Diagrama de Software TranscripApp*



***Nota*:** Elaboración propia

Descripción Workflow de TranscripApp

- **Paso 1:** Desde la **Interfaz de usuario (UI)**, se deberá seleccionar el archivo de audio el será objeto de transcripción, esta acción dispara la validación del ***Control de Permisos***.

- **Paso 2:** la validación de privilegios requeridos son READ_EXTERNAL_STORAGE, el cual es mandatorio para que TranscripApp funcione.

- **Paso 3:** Posterior a la validación del paso 2 el bloque de **Transcripción** realiza la integración del prototipo con OpenAI.

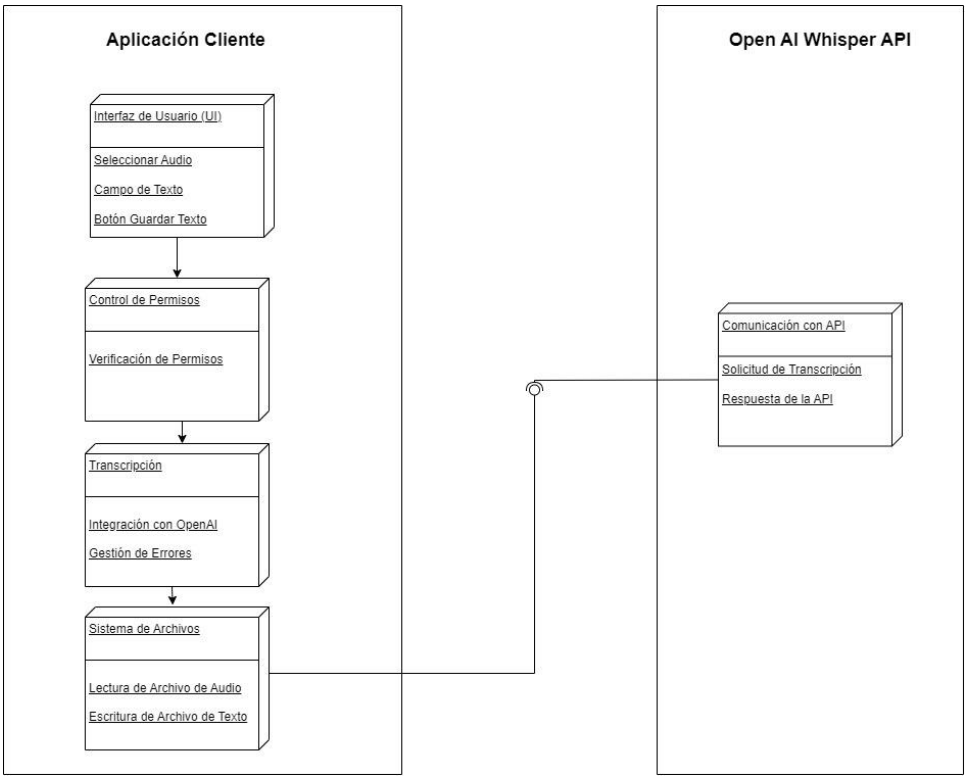
- **Paso 4:** En el **Sistema de Archivos** se realiza una validación del archivo y lectura del archivo de audio seleccionado en el bloque inicial **Interfaz de usuario (UI)** esto como preparación del request.

- **Paso 5:** Si el paso 4 fue completado satisfactoriamente se realiza **Comunicación con API** y se realiza el request solicitando la transcripción del audio usando OpenAI *Whisper*.

- **Paso 6:** Posterior a la transcripción se realiza el response con la data es decir la transcripción del archivo de audio este es presentado en la **Interfaz de usuario (UI).**

- **Paso 7: S**i el usuario lo requiere puede guardar las transcripción es decir el texto para su consulta o uso posterior.

Figura 4 *Diagrama de bloques TranscripApp.*



***Nota*:** Elaboración propia

Descripción del Desarrollo TranscripApp

Teniendo presente el corto espacio para el desarrollo del proyecto decimos realizar el desarrollo e implantación del prototipo de aplicación TranscripApp en lenguaje Kotlin usando como IDE el programa Android Studio, el cual ofrece una curva de aprendizaje moderada en comparación con otros lenguajes de programación para aplicaciones móviles como Xamarin o Java.

En el desarrollo se implementó el uso o consumo de la inteligencia artificial de OpenAI llamada Whisper, la cual ofrece una alternativa robusta en el ámbito de reconocimiento automático de voz (ASR), Esta inteligencia artificial permite su interacción a través de una API de muy bajo costo de 1 centavo USD por petición lo cual hace que TranscripApp sea una herramienta posible de financiar a través de publicidad embebida.

Se definieron 5 bloques de los cuales funcionan de manera secuencial disparando un bloque a la vez dependiendo el estado del proceso principal de cada bloque anterior, iniciando en el bloque de la interfaz gráfica al seleccionar el archivo de audio objeto de transcripción y finalizando en este mismo bloque presentando el texto de la transcripción, todo el proceso realizado a través del bloque “Try” que encapsula el proceso de transcripción y si se produce una excepción se captura a través del bloque catch, finalmente se registra en el Log.E y se muestra el mensaje en la aplicación registrado a través de “logs” en archivos de texto plano que ofrecen detalle de su ejecución en un nivel de detalle mínimo para no comprometer el rendimiento de la aplicación, es importante recalcar que todo este procedimiento es transparente para el usuario y se realiza de manera no visual para el usuario.

Resultados

En la presente sección se evidencian los resultados de las pruebas ejecutadas a los 14 archivos de audio digital creados por familiares, amigos e integrantes del grupo GIS 03A en diferentes formatos MP3, WAV y Mp4, estos archivos de audio se crearon a partir de la lectura y grabación de un fragmento de la

obra literaria "*Cien Años de Soledad*" autoría de Gabriel Garcia Márquez, "Muchos años después, frente al pelotón de fusilamiento, el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre 10 llevó a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y cañabrava construidas a la orilla de un río de aguas diáfanas que se precipitaban por un lecho de piedras pulidas, blancas y enormes como huevos prehistóricos. EI mundo era tan reciente, que muchas cosas carecían de nombre, y para mencionarlas había que señalarlas con el dedo." (l, Gabriel García, Cien años de soledad).

Estos archivos de audio grabados sirvieron de insumo para pruebas de funcionalidad en cada uno de los componentes de TranscripApp; de igual manera se podrá observar a través del diagrama de bloques la interacción entre los componentes del prototipo de aplicación en Android TranscripApp.

Los indicadores establecidos para alcanzar un estándar mínimo en desarrollo del prototipo y cumplimiento de los objetivos son:

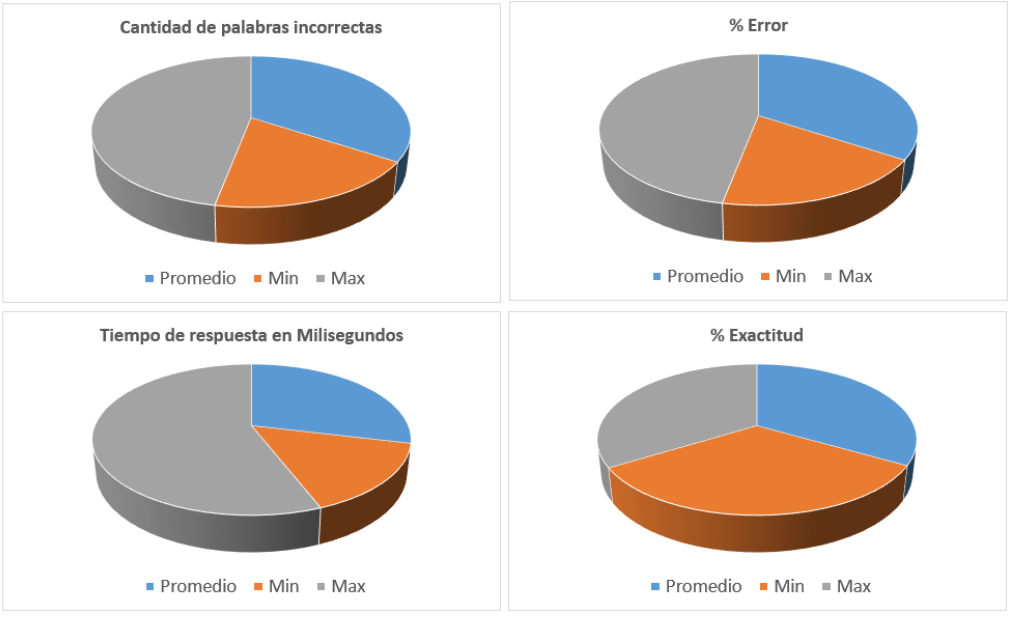
1. **Cantidad de palabras incorrectas:** Indicador basado en el conteo de palabras erradas en proceso de transcripción.
2. **Tiempo de respuesta:** Indicador basado en el conteo de tiempo total del proceso de transcripción medido en Milisegundos.
3. **Porcentaje de Error:** Indicador basado en conteo total de palabras detectadas en el archivo de audio en comparación con el total de palabras incorrectas detectadas durante la transcripción el cual se expresa en un valor porcentual.
4. **Porcentaje de exactitud:** Indicador basado en conteo total de palabras detectadas en el archivo de audio en comparación con el total de palabras correctamente detectadas durante la transcripción el cual se expresa en un valor porcentual.

Tabla 2 *Resultados de pruebas TranscripApp.*

Subject	Cantidad de palabras incorrectas	Tiempo de respuesta en Milisegundos	% Error	% Exactitud	palabras
1	3	2676	2.55	97.45	85
2	3	3298	2.55	97.45	85
3	5	3032	4.25	95.75	85
4	2	8429	1.7	98.3	85
5	5	2309	4.25	95.75	85
6	4	8203	3.4	96.6	85
7	4	8079	3.4	96.6	85
8	4	8066	3.4	96.6	85
9	4	2945	3.4	96.6	85
10	3	2524	2.55	97.45	85
11	4	2803	3.4	96.6	85
12	4	2651	3.4	96.6	85
13	3	2896	2.55	97.45	85
14	3	2706	2.55	97.45	85
Average	3.64	4329.79	3.10	96.90	
Min	2	2309	1.7	95.75	
Max	5	8429	4.25	98.3	

***Nota*:** Elaboración propia.

Figura 5 *KPI's resultados de pruebas TranscripApp*



***Nota*:** Elaboración propia

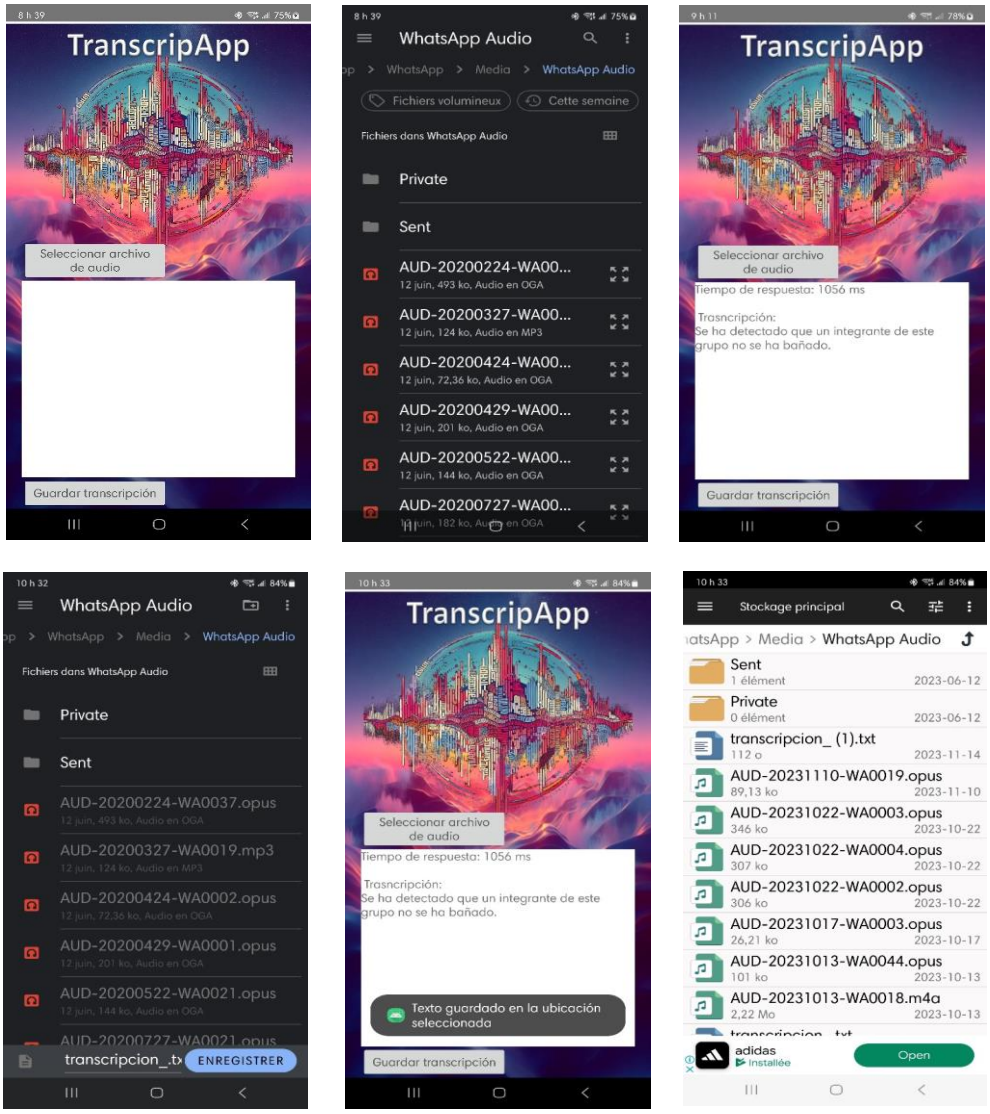
Se realizaron 6 rondas de transcripción a cada uno de los 14 archivos de audio esto con el objetivo de validar resultados entre las transcripciones, reduciendo las excepciones o errores en el proceso de transcripción de

los audios, adicionalmente nuestro prototipo TranscripApp entrena la API de Whisper para el uso del idioma español cada vez que se ejecuta una transcripción.

Interfaz Gráfica de usuario

Se diseño una interfaz grafica sencilla y sobria que ofrece una manera fácil para comprender el uso de la única y principal funcionalidad de transcripción de audio a texto.

Figura 6 *Interfaz Gráfica TranscripApp*



***Nota*:** Elaboración propia

Discusión Aplicaciones similares

En la tienda de aplicaciones Google Play Store se pueden encontrar varias aplicaciones para transcripción de audio a texto y subtitulado pero, haciendo una exhaustiva verificación de comentarios de usuarios de estas se evidenció que solo dos aplicaciones cumplen parcialmente la función principal de transcripción de audio como característica principal, estas aplicaciones son: *Transcripción Instantánea de Google*, o *Ava: Transcriptions & Captions de Ava Accessibility*, pero su principal mercado está dirigido al uso en el idioma inglés, lo cual incentiva el desarrollo de nuestro prototipo de aplicación TranscripApp, en el siguiente cuadro comparativo se podrá en evidenciar sus principales características.

Tabla 3 *Cuadro comparativo aplicaciones de transcripción en Google Play Vs TranscripApp.*

	Ava: Transcriptions & Captions	Transcripción Instantánea	TranscripApp
Desarrollador	Ava Accessibility	Google	Grupo GIIS-03A UCompensar
Calificacion de Google Play	3,4	4.2	-
Facilidad de uso según usuarios	Normal	Medio	Facil
Precision según cada fabricante	70% en Español	85% en Español	97% en Español
Funciona en Español	No	No	Si
Compatibilidad según Google Play	Android 12 en adelante	Android 11 en adelante	Android 4.4 en adelante
Costo en Google Play	Gratis version base por 5 horas	Gratis version base	Gratis *incluye publicidad
Opiniones de los usuarios	3,3	3,6	-
Última Actualización	29 ago 2023	31 ago 2023	20-Oct-23
Soporte *Condicionado	Si	Si	Si
Licencia de uso	Limitada	Limitada	Limitada
Recogen datos de uso y otras Apps	Si	Si	No
Contiene anuncios	Si	No	Si

***Nota*:** Elaboración propia.

Referencias

1. OMS. (2023). Sordera y pérdida de la audición. Organización Mundial de la Salud. https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss
2. OPS. (2023). Muchos países carecen de la capacidad para prevenir y tratar la pérdida de audición. Organización Panamericana de la Salud. https://www3.paho.org/hq/index.php?Itemid=0&id=9345%3A2014-many-countries-lack-capacity-prevent-treat-hearing-loss&lang=es&option=com_content&view=article
3. International Speech Communication Association. (2023). Newsroom. https://www.isca-speech.org/iscaweb/index.php/newsroom
4. Eduardo Esteban Pérez León. (2023). Discapacidad auditiva y pedagogías emergentes: Desafíos y retos de un modelo de educación mediado por las tecnologías digitales. Instituto Técnico Guaimaral. https://revista.redipe.org/index.php/1/article/download/1824/1748/2937
5. Congreso de la República de Colombia. (2005). Ley 982 de 2005. Función pública. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=17283#:~:text=Es%20la%20que%20reconoce%20que,oral%20en%20los%20casos%20en
6. MinTIC. (2015). MinTIC lanza aplicación móvil para personas con discapacidad auditiva. MinTIC. https://www.mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/13849:MinTIC-lanza-aplicacion-movil-para-personas-con-discapacidad-auditiva
7. DANE. (2022). Encuesta Nacional de Calidad de Vida -ECV- 2022. DANE Colombia. https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/calidad-de-vida-ecv/encuesta-nacional-de-calidad-de-vida-ecv-2022
8. Min Salud. (2013). Ley Estatutaria 1618 De 2013. Ministerio de Salud Colombia. https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/documento-balance-1618-2013-240517.pdf

9. INSOR. (2023). Guía para la producción de contenidos audiovisuales con ajustes razonables para la población sorda colombiana. www.insor.gov.co. https://www.insor.gov.co/home/descargar/Guia-para-la-produccion-de-contenidos-audiovisuales-con-ajustes-razonables-para-la-poblacion-sorda-colombiana.pdf
10. NIDCD . (2017). Subtítulos para las personas sordas o con problemas de audición. www.nidcd.gov.es. https://www.nidcd.nih.gov/es/espanol/subtitulos
11. Ava Accessibility. (2022). Ava: Transcriptions & Captions. https://www.ava.me/. https://play.google.com/store/apps/details?id=com.transcense.ava_beta&hl=es&gl=US
12. Google. (2023). Transcripción instantánea. . https://play.google.com/store/apps/details?id=com.google.audio.hearing.visualization.accessibility.scribe&hl=es_AR
13. García, G. [Ed.]. (1967), Cien años de soledad, Colombia: Editorial planeta.
14. Etchegoyhen, T.; Arzelus, H.; Gete, H.; Alvarez, A.; Torre, I.G.; Martín-Doñas, J.M.; González-Docasal, A.; Fernandez, (2022). Cascade or Direct Speech Translation? A Case Study. MDPI. https://doi.org/10.3390/app12031097
15. Thomas Reitmaier, Electra Wallington, Dani Kalarikalayil Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. 2022. Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers. In Proceedings of the (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 299, 1–17. https://doi.org/10.1145/3491102.3517639
16. Rascon, C.; Fuentes-Pineda, G. (2023). Target Selection Strategies for Demucs-Based Speech Enhancement. MDPI. https://doi.org/10.3390/app13137820
17. Catherine Jewell, Hadeel Ayoub. (2019). BrightSign: el guante inteligente que da voz a quienes no pueden hablar. OMPI. https://www.wipo.int/wipo_magazine/es/2019/05/article_0005.html