UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO
CARRERA DE BIOMEDICINA

DESARROLLO E IMPLEMENTACIÓN DE UN SISTEMA DE LECTURA ASISTIDA MEDIANTE VOZ PARA PERSONAS INVIDENTES CON TÉCNICAS DE VISIÓN ARTIFICIAL E INTELIGENCIA ARTIFICIAL.

Trabajo de titulación previo a la obtención del Título de:
INGENIERO BIOMÉDICO

AUTOR: FERNANDO ESTEBAN PAREDES PAREDES

TUTOR: ING. LUIS GEOVANNY ROMERO MEJÍA

Quito-Ecuador

2025

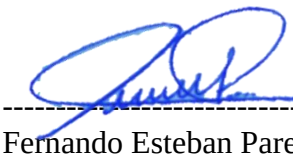
CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Fernando Esteban Paredes Paredes con documento de identificación N° 1723069140 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 17 de Febrero del año 2025

Atentamente,



Fernando Esteban Paredes Paredes
1723069140

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Fernando Esteban Paredes Paredes con documento de identificación No. 1723069140, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto Técnico: "DESARROLLO E IMPLEMENTACIÓN DE UN SISTEMA DE LECTURA ASISTIDA MEDIANTE VOZ PARA PERSONAS INVIDENTES CON TÉCNICAS DE VISIÓN ARTIFICIAL E INTELIGENCIA ARTIFICIAL", el cual ha sido desarrollado para optar por el título de: Ingeniero en Biomedicina en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 17 de Febrero del año 2025

Atentamente,

Fernando Esteban Paredes Paredes

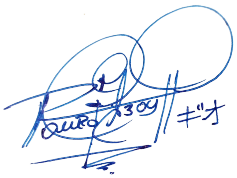
1723069140

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Luis Geovanny Romero Mejía con documento de identificación N° 1714731203, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: "DESARROLLO E IMPLEMENTACIÓN DE UN SISTEMA DE LECTURA ASISTIDA MEDIANTE VOZ PARA PERSONAS INVIDENTES CON TÉCNICAS DE VISIÓN ARTIFICIAL E INTELIGENCIA ARTIFICIAL", realizado por Fernando Esteban Paredes Paredes con documento de identificación N° 1723069140, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto Técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 17 de Febrero del año 2025

Atentamente,



Luis Geovanny Romero Mejía

1714731203

Dedicatoria

A mis padres,

por su amor incondicional,

por ser mi fuente de inspiración

y por enseñarme que con esfuerzo y dedicación

todo es posible.

Su apoyo constante y sus palabras de aliento

me han guiado en cada paso de este camino.

Este logro es tan suyo como mío.

Resumen

Este proyecto aborda el diseño, desarrollo e implementación de un sistema de lectura asistida mediante voz, especialmente orientado a personas con discapacidad visual. Utilizando técnicas avanzadas de visión artificial e inteligencia artificial, el sistema convierte de manera precisa y eficiente texto impreso en formato de audio, facilitando el acceso inclusivo a la información en entornos educativos, laborales y cotidianos.

El trabajo incluyó la selección y configuración de componentes de hardware, como una Raspberry Pi 4, una cámara Logitech C920s y altavoces Genius SP-U115, junto con herramientas de software como el modelo EAST para detección de texto, Tesseract OCR para reconocimiento óptico de caracteres y gTTS para la síntesis de voz. Además, se integraron mecanismos de interacción, como comandos mediante teclado y retroalimentación auditiva, para garantizar una experiencia accesible y autónoma.

Las pruebas realizadas en diversos escenarios de iluminación confirmaron un rendimiento destacado del sistema, alcanzando precisiones de hasta un 95 % bajo condiciones óptimas. Se evaluaron también tiempos de procesamiento y claridad del audio generado, demostrando que el dispositivo es viable para un uso en tiempo real. Este proyecto representa un avance significativo hacia la inclusión tecnológica, mejorando la calidad de vida de las personas con discapacidad visual y sentando las bases para futuras mejoras y expansiones.

Palabras clave: visión artificial, inteligencia artificial, OCR, lectura asistida, accesibilidad.

Abstract

This project addresses the design, development, and implementation of a voice-assisted reading system specifically aimed at visually impaired individuals. By leveraging advanced techniques in computer vision and artificial intelligence, the system accurately and efficiently converts printed text into audio format, enabling inclusive access to information in educational, workplace, and daily environments.

The work involved selecting and configuring hardware components such as a Raspberry Pi 4, a Logitech C920s camera, and Genius SP-U115 speakers, along with software tools like the EAST model for text detection, Tesseract OCR for optical character recognition, and gTTS for voice synthesis. Additionally, interaction mechanisms, including keyboard commands and auditory feedback, were integrated to ensure an accessible and autonomous user experience.

Tests conducted in various lighting scenarios confirmed the system's remarkable performance, achieving accuracies of up to 95 % under optimal conditions. Processing times and audio clarity were also evaluated, demonstrating the device's feasibility for real-time use. This project represents a significant step towards technological inclusion, enhancing the quality of life for visually impaired individuals and laying the groundwork for future improvements and expansions.

Keywords: computer vision, artificial intelligence, OCR, assisted reading, accessibility.

Índice

Lista de Tablas

Lista de Figuras

1. Introducción

1.1. Antecedentes

A lo largo de la historia, el acceso a la información escrita ha representado un desafío particularmente complejo para las personas con discapacidad visual, limitando sus oportunidades educativas, laborales y su plena integración social. De acuerdo con datos de la Organización Mundial de la Salud (2019), más de 2.200 millones de personas presentan alguna forma de deficiencia visual o ceguera, de las cuales al menos 1.000 millones podrían haberse evitado o aún no han recibido tratamiento adecuado. Entre las principales causas que originan esta situación se encuentran las cataratas, el glaucoma y los errores de refracción no corregidos.

En América Latina, aproximadamente 26 millones de personas sufren discapacidad visual, y cerca de 3 millones son ciegas (Sociedad Panamericana de Retinopatía del Prematuro, 2020). En el caso del Ecuador, según el Ministerio de Salud Pública de Ecuador (2022), 73.771 individuos viven con algún tipo de discapacidad visual, siendo las cataratas y los errores de refracción no corregidos las causas más frecuentes (Organización Panamericana de la Salud, s.f.). Esta realidad se acentúa en áreas rurales e indígenas, donde el limitado acceso a servicios oftalmológicos de calidad impacta negativamente en la calidad de vida y las oportunidades de las personas, incrementando también el riesgo de depresión y ansiedad.

Comparado con otros países de la región, la situación en Ecuador guarda similitudes con las de Brasil o México, en donde las cataratas siguen siendo una causa importante de ceguera (The international agency for the prevention of blindness, 2021). Estas coincidencias subrayan la necesidad de soluciones tecnológicas accesibles y eficientes, capaces de beneficiar a poblaciones que enfrentan constantes obstáculos para acceder a la información escrita.

Desde el siglo XIX, el sistema Braille, creado por Louis Braille, marcó un antes y un después en el acceso a la lectura para las personas ciegas, permitiendo que interactúen con textos mediante un código táctil en relieve (Antonacopoulos & Bridson, 2004). Con el paso del tiempo, el siglo XX

trajo consigo dispositivos electrónicos, lectores de pantalla y sistemas de reconocimiento óptico de caracteres (OCR) que digitalizaron el texto, facilitando su uso en formatos accesibles. Aun así, en sus inicios, estas tecnologías enfrentaron limitaciones técnicas, como requerir iluminación controlada y tener dificultades para interpretar textos en entornos cotidianos complejos (Yellapu et al., 2022).

En la actualidad, la integración de visión artificial e inteligencia artificial ha permitido superar muchas de estas barreras. La visión artificial, que dota a los sistemas de la capacidad de procesar y analizar imágenes del mundo real, ha incrementado la versatilidad del reconocimiento de texto. Por su parte, los algoritmos de aprendizaje profundo han mejorado notablemente la precisión y robustez de los sistemas OCR, habilitándolos para reconocer patrones más complejos y adaptarse a situaciones diversas (Huynh et al., 2023).

2. Problemática

A pesar de las mejoras tecnológicas de las últimas décadas, el acceso a sistemas asistidos que ofrezcan información escrita de manera fluida y autónoma sigue siendo una barrera relevante para las personas con discapacidad visual. Esta situación, que afecta a su desempeño educativo, sus oportunidades laborales y su participación en la vida social, se ve agravada por entornos con recursos limitados. Aunque herramientas como el Braille, los lectores de pantalla y algunos sistemas OCR han facilitado parcialmente el acceso a la información, su efectividad en condiciones reales dista de ser óptima.

En el contexto latinoamericano, y muy especialmente en el Ecuador, la discapacidad visual afecta a más de 55.000 personas, situándolas ante importantes limitaciones funcionales (Consejo Nacional para la Igualdad de Discapacidades, 2024). En zonas rurales, donde el acceso a servicios oftalmológicos y tecnologías de apoyo es muy restringido, la necesidad de soluciones tecnológicas adaptadas a la realidad cotidiana es aún más urgente. Sin embargo, las alternativas actuales, como dispositivos OCR o lectores de pantalla, no logran ofrecer una experiencia completa e intuitiva en

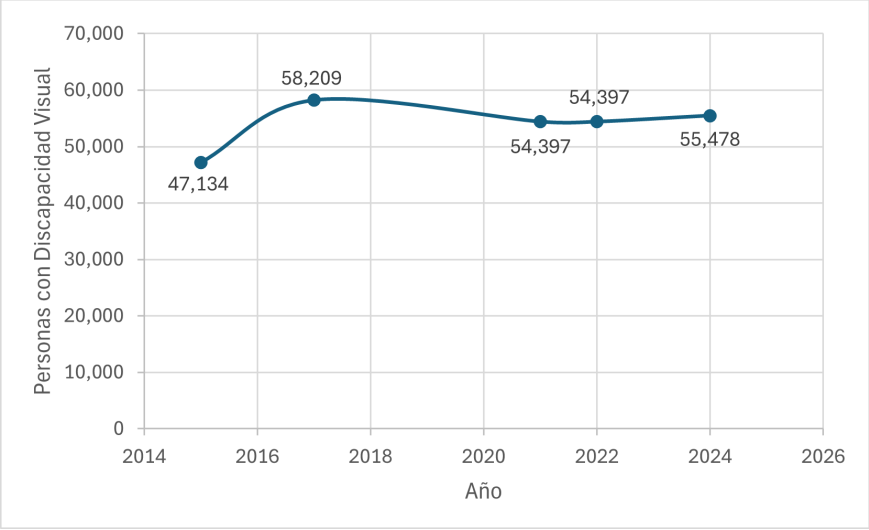
condiciones variables de iluminación, tipografía o disposición espacial del texto.

La incorporación de técnicas de visión artificial e inteligencia artificial a un sistema de lectura asistida mediante voz se plantea aquí como una alternativa eficaz y necesaria. Si bien los dispositivos modernos pueden ofrecer un reconocimiento de texto bastante preciso en entornos controlados, persisten desafíos cuando se trata de escenarios complejos, textos desalineados o condiciones lumínicas adversas (Sarwar et al., 2022).

La Figura 1 muestra el incremento en los casos de discapacidad visual en Ecuador entre 2015 y 2024, evidenciando una tendencia ascendente. Este contexto reclama soluciones tecnológicas que no solo mejoren la precisión en el reconocimiento, sino que resulten accesibles, asequibles y adaptadas a las necesidades reales de las personas. En definitiva, se requiere un sistema que integre reconocimiento de texto, conversión a audio y la posibilidad de operar de manera confiable y autónoma, incluso en entornos cotidianos con recursos limitados.

Figura 1

Discapacidad Visual en Ecuador (2015-2024).



Nota: Datos obtenidos del Ministerio de Salud Pública de Ecuador (2022) y del Consejo Nacional para la Igualdad de Discapacidades (2024) .

2.1. Justificación

La necesidad de un sistema de lectura asistida basado en inteligencia artificial y visión artificial surge de la urgencia social de ofrecer herramientas que permitan a las personas con discapacidad visual acceder a la información de manera más autónoma. En una sociedad donde la palabra escrita es fundamental para la educación, el trabajo y la vida cotidiana, no contar con soluciones tecnológicas adaptadas limita gravemente el desarrollo personal, profesional y social de quienes viven con una deficiencia visual (Organización Mundial de la Salud, 2019).

Este proyecto busca proporcionar una respuesta asequible y práctica, utilizando componentes disponibles en el mercado, como la Raspberry Pi, y algoritmos de código abierto, lo que abarata costos y facilita la implementación en diferentes contextos. De este modo, la tecnología no queda restringida a entornos privilegiados, sino que puede llegar a comunidades con menos recursos, promoviendo así la equidad en el acceso a la información.

La inteligencia artificial aplicada al OCR brinda una precisión y adaptabilidad superiores, ya que mediante algoritmos de aprendizaje profundo se pueden reconocer textos complejos en condiciones adversas. Asimismo, integrar un módulo de conversión de texto a voz optimizado para dispositivos con recursos limitados permitirá una respuesta ágil y fiable sin requerir hardware sofisticado. En conjunto, esta solución puede mejorar la autonomía de las personas con discapacidad visual, incrementando su participación activa en la educación, el empleo y la vida social, al tiempo que se convierte en una herramienta más fácil de adquirir y utilizar en diversos contextos.

2.2. Objetivos

2.2.1. Objetivo General

Desarrollar e implementar un sistema de lectura asistida mediante voz para personas invidentes, utilizando técnicas avanzadas de visión artificial e inteligencia artificial para la conversión precisa y eficiente de texto impreso a formato de audio.

2.2.2. Objetivos Específicos

- Identificar los requisitos técnicos y funcionales para el sistema de lectura asistida, incluyendo las especificaciones de hardware y software necesarios para la captura de imágenes y procesamiento de texto, con la revisión exhaustiva del estado del arte.
- Analizar las técnicas de visión artificial e inteligencia artificial más adecuadas para la captura y reconocimiento de texto, utilizando técnicas y algoritmos ORC (Algoritmos de Reconocimiento Óptico de Caracteres) y procesamiento de imágenes.
- Evaluar el rendimiento del prototipo en términos de precisión, velocidad y adaptabilidad bajo diferentes condiciones de iluminación y tipos de texto realizando pruebas exhaustivas en distintos escenarios.

3. Fundamento Teórico

3.1. Discapacidad Visual y Acceso a la Información

La discapacidad visual se refiere a una disminución significativa de la capacidad de ver que no puede corregirse por completo con lentes, medicamentos o cirugía. Esta condición abarca desde una reducción leve de la visión hasta la ceguera total. De acuerdo con la Organización Mundial de la Salud (s.f.), la función visual se clasifica en cuatro categorías principales:

- Visión normal: Agudeza visual de 20/20 o mejor.
- Discapacidad visual moderada: Agudeza visual entre 20/70 y 20/160.
- Discapacidad visual grave: Agudeza visual entre 20/200 y 20/400.
- Ceguera: Agudeza visual peor que 20/400 o campo visual menor de 10 grados.

Estas categorías facilitan la identificación del nivel de adaptación y asistencia que una persona puede requerir en su vida cotidiana.

3.1.1. Impacto de la Discapacidad Visual en el Acceso a la Información

La discapacidad visual repercute de manera notable en la capacidad de las personas para acceder a la información, afectando así múltiples dimensiones de su vida:

- **Educación:** Tal como indican Curiel Centenero y et al. (2018), los estudiantes con discapacidad visual se enfrentan a desafíos al acceder a materiales educativos estándar, lo que dificulta su aprendizaje y su pleno desarrollo académico. La ausencia de recursos adecuados (libros en braille, audiolibros, o tecnologías asistivas) limita su integración en el entorno educativo.

- **Empleo:** Según Pallero (2008), la falta de adaptaciones en el lugar de trabajo y la escasez de tecnologías accesibles obstaculizan las oportunidades laborales. Esta situación reduce las posibilidades de crecimiento profesional y restringe el acceso a un empleo pleno y significativo.

- **Vida cotidiana:** De igual manera, Pallero (2008) destaca que las dificultades para acceder a información escrita en espacios públicos (señalética, pantallas, instrucciones de productos) disminuyen la autonomía diaria. Esto puede traducirse en un aislamiento social y un impacto negativo en la calidad de vida.

3.2. Tecnologías de Asistencia para Personas con Discapacidad Visual

3.2.1. Evolución de los Dispositivos de Asistencia

La evolución de las tecnologías de asistencia ha sido clave para mejorar la calidad de vida de las personas con discapacidad visual. Desde el siglo XIX, el sistema Braille, desarrollado por Louis Braille, ha representado un pilar en la alfabetización de personas ciegas al codificar letras y números en puntos en relieve (Organización Mundial de la Salud, 2019).

Con la llegada de la era digital, surgieron dispositivos más avanzados, entre ellos la tecnología de Reconocimiento Óptico de Caracteres (OCR) y los lectores de pantalla, que permiten a los usuarios interactuar con computadoras y dispositivos móviles. Estas herramientas convierten el

contenido visual en audio o en salidas braille, contribuyendo al acceso a la información digital. No obstante, como señalan Leporini y Paternò (2004), aún persisten limitaciones, entre ellas la dependencia de condiciones de iluminación óptimas o la dificultad para interpretar entornos no estructurados. Además, la experiencia auditiva puede diferir significativamente de la experiencia visual, lo que plantea desafíos para la comprensión integral del contenido.

3.2.2. Reconocimiento Óptico de Caracteres (OCR)

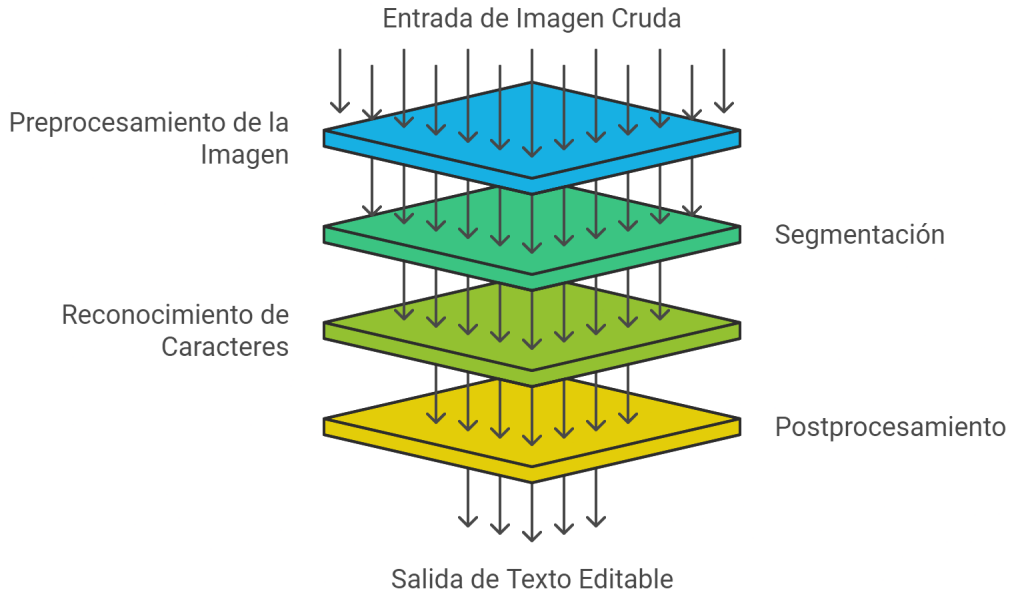
El Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés) constituye una tecnología esencial para transformar texto impreso o manuscrito en datos digitales editables. Este proceso resulta fundamental para mejorar la accesibilidad a la información de las personas con discapacidad visual, ya que convierte documentos físicos en formatos legibles por lectores de pantalla o dispositivos braille.

Los primeros sistemas de OCR aparecieron en la década de 1950 y se enfocaban en reconocer caracteres impresos en fuentes específicas. Con el tiempo, y gracias a los avances en inteligencia artificial y aprendizaje automático, estas herramientas han evolucionado para abarcar una amplia gama de fuentes y estilos, incluyendo texto manuscrito. Tal como señalan Cheriet et al. (2007), los sistemas modernos de OCR han incrementado su precisión y versatilidad en diversos idiomas y contextos. De igual manera, Hamad y Kaya (2016) resaltan que la exactitud del OCR depende de factores como la calidad de la imagen, el tipo de fuente y la complejidad del diseño.

Como se muestra en la Figura 2, el flujo del proceso OCR abarca desde la captura de la imagen hasta la generación del texto digitalizado, pasando por las etapas de preprocesamiento, segmentación, reconocimiento y postprocesamiento. Esta secuencia asegura una conversión eficiente y precisa del contenido impreso a un formato accesible.

Figura 2

Flujo del proceso de Reconocimiento Óptico de Caracteres (OCR).



Nota: Elaboración propia.

Como destacan Chaudhuri et al. (2017), los sistemas OCR más recientes incorporan técnicas de soft computing, lo que mejora la precisión en una variedad de lenguajes y condiciones.

Las ventajas del OCR para las personas con discapacidad visual son múltiples:

- **Acceso a la Información:** Permite la conversión de materiales impresos a formatos digitales accesibles, facilitando su lectura mediante tecnologías asistivas.
- **Autonomía:** Reduce la dependencia de terceros para la lectura de documentos físicos.
- **Inclusión Educativa y Laboral:** Facilita la participación en entornos académicos y profesionales al proporcionar acceso a materiales y recursos previamente inaccesibles.

Fichten et al. (2009) enfatizan que la accesibilidad de las tecnologías de la información y el e-learning es crucial para la inclusión de estudiantes con discapacidades visuales en la educación superior.

3.2.3. Lectura Asistida mediante Voz

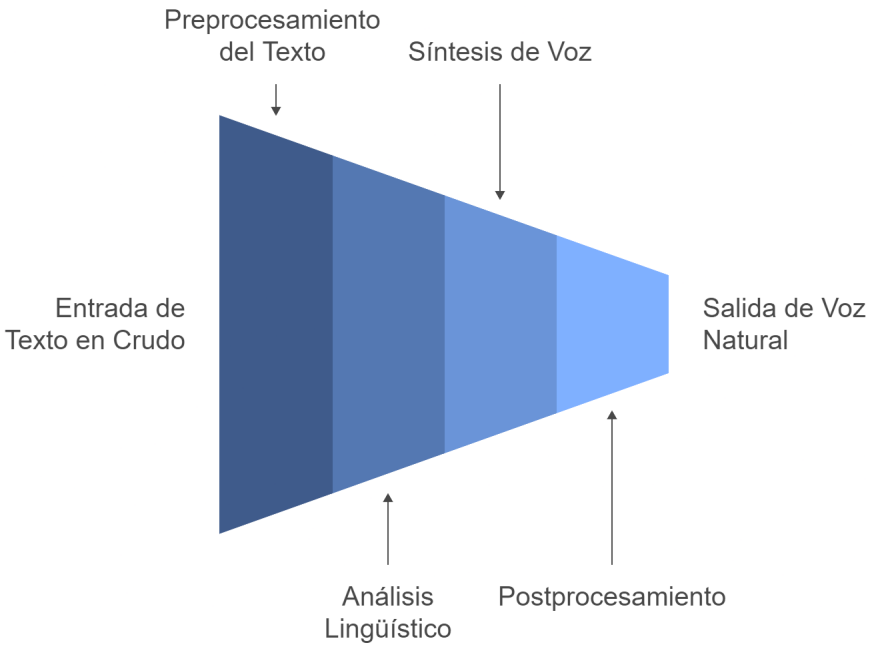
La lectura asistida mediante voz es una tecnología que convierte texto escrito en audio, facilitando el acceso a la información para personas con discapacidad visual. Este proceso se basa en sistemas de conversión de texto a voz (TTS, por sus siglas en inglés) que analizan el texto y lo transforman en lenguaje hablado. Según Rasinski y Young (2014), la lectura asistida actúa como un puente entre la fluidez lectora y la comprensión, mejorando la experiencia de lectura para los usuarios.

Inicialmente este sistema de conversión texto a voz producía voces robóticas y monótonas, sin la fluidez y ni acentuación que tendría una persona normal. Sin embargo, con los avances en inteligencia artificial y aprendizaje profundo, se han desarrollado voces más naturales y expresivas. Modelos como WaveNet de Google han sido fundamentales en esta transición, permitiendo una síntesis de voz que se asemeja más al habla humana natural (van den Oord et al., 2016).

La Figura 3, ilustra el flujo del proceso TTS que abarca desde el preprocesamiento del texto hasta el postprocesamiento del audio generado. Esta orden establece una conversión eficiente y precisa del contenido textual a un formato de audio comprensible.

Figura 3

Flujo del proceso de Text-to-Speech (TTS).



Nota: Elaboración propia.

Hoy en día, existen múltiples dispositivos y aplicaciones que incorporan tecnología TTS para brindar autonomía a personas con discapacidad visual. Por ejemplo, OrCam MyEye se integra en las gafas y lee en voz alta textos impresos y digitales, además de reconocer rostros y productos, lo que facilita enormemente las actividades cotidianas (OrCam Staff, 2022). De manera similar, la aplicación móvil Seeing AI de Microsoft describe el entorno, lee documentos y reconoce objetos, aumentando la independencia de sus usuarios (Microsoft Corporation, s.f.).

Algunos estudios presentan iniciativas similares. Por ejemplo, Ani et al. (2017) describen un sistema basado en OCR y TTS, utilizando herramientas como Tesseract OCR y eSpeak en una plataforma Raspberry Pi, para leer texto en voz alta a personas con discapacidad visual. Esta solución permite que los usuarios escuchen el contenido en tiempo real, incrementando su autonomía y mejorando el acceso a la información.

Además, investigaciones han demostrado que la lectura asistida mediante audio puede mejorar las tasas de lectura y la comprensión, incluso en aprendices principiantes. Un estudio de Chang y Millett (2015) evidencia que el uso de audio en la lectura extensiva asistida beneficia a los estudiantes, incrementando su velocidad y comprensión lectora.

En el ámbito de la inteligencia artificial aplicada a la asistencia visual, M. A. Khan et al. (2020) desarrollaron una ayuda para personas con ceguera total, integrando asistencia de lectura y navegación electrónica. Este tipo de enfoques combinan varias tecnologías, ofreciendo soluciones más integrales y impactando positivamente en la calidad de vida de los usuarios.

La lectura asistida mediante voz ofrece múltiples beneficios:

- **Acceso a la información:** Permite a las personas con discapacidad visual disfrutar de textos impresos y digitales a través del audio.
- **Autonomía:** Disminuye la dependencia de terceros en la lectura de documentos y señalética.
- **Inclusión educativa y laboral:** Facilita la participación en entornos académicos y profesionales, proporcionando acceso a materiales que antes resultaban inaccesibles.

3.3. Inteligencia Artificial y Visión Artificial en Sistemas de Asistencia

3.3.1. Fundamentos de la Visión Artificial

La inteligencia artificial (IA) y la visión artificial (VA) han transformado los sistemas de asistencia, especialmente para personas con discapacidad visual, al posibilitar que los dispositivos comprendan e interpreten información visual del entorno. Esto les permite ofrecer respuestas útiles y accesibles, mejorando su integración social y su autonomía en la vida diaria.

La visión artificial, una rama de la IA, emplea algoritmos avanzados para analizar imágenes y extraer información significativa. Con los constantes avances en aprendizaje automático y redes neuronales, la visión artificial ha ganado precisión y capacidad de adaptación, siendo clave para la detección de objetos, reconocimiento facial y análisis de escenas. De acuerdo con Elyan et al.

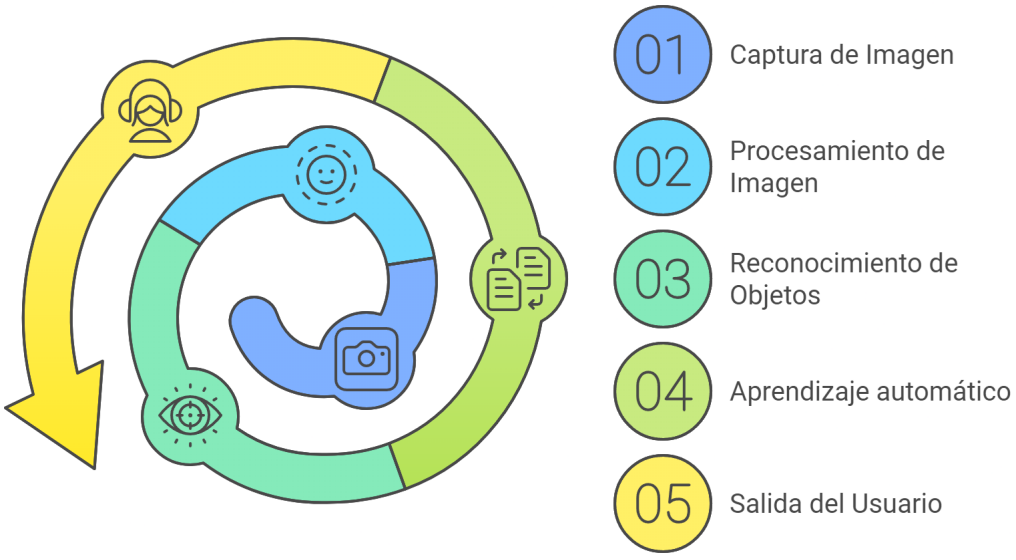
(2022), el uso de técnicas de aprendizaje profundo ha impulsado el desarrollo de sistemas más precisos en diversos campos, desde la medicina hasta la asistencia visual.

Asimismo, el uso de modelos de IA a gran escala ha fortalecido la visión artificial en aplicaciones que requieren alta capacidad de procesamiento. Por ejemplo, Lin (2022) destaca que el crecimiento de estos modelos ha incrementado la potencia y la adaptabilidad de la visión artificial, elementos cruciales para tareas de asistencia donde se necesita procesamiento en tiempo real.

Los sistemas de visión artificial para asistencia siguen un flujo básico como se visualiza en la Figura 4: la imagen se captura, se procesa y analiza a través de algoritmos, y finalmente se genera una respuesta comprensible para el usuario, ya sea en formato auditivo u otra modalidad sensorial. Tal como explica A. I. Khan y Al-Habsi (2020), el aprendizaje automático permite que estos sistemas se adapten a diferentes tipos de imágenes y patrones visuales, lo que resulta esencial para una interpretación precisa en una amplia variedad de escenarios.

Figura 4

Arquitectura del sistema de visión artificial.



Nota: Elaboración propia.

Hoy en día, la visión artificial se integra en múltiples dispositivos y aplicaciones de asistencia.

Por ejemplo, retomando la aplicación Seeing AI de Microsoft convierte las imágenes del entorno en descripciones auditivas, ofreciendo información en tiempo real (Microsoft Corporation, s.f.), mientras que OrCam MyEye combina visión artificial y síntesis de voz para leer textos, reconocer rostros y describir escenas (OrCam Staff, 2022). Estas soluciones promueven la independencia de los usuarios y mejoran su experiencia diaria.

No obstante, la visión artificial enfrenta desafíos: la precisión puede verse afectada por condiciones de iluminación variables y contextos complejos. Aunque Lin (2022) señala que los modelos de IA a gran escala resultan prometedores, su implementación en dispositivos portátiles implica gestionar recursos limitados. De igual manera, Kashyap y Kumar (2019) analiza las dificultades técnicas al aplicar aprendizaje automático en visión artificial, particularmente en sistemas que exigen alta precisión y velocidad. Además, Suma (2019) subraya que la interacción humano-máquina a través de visión artificial aún debe superar retos significativos para alcanzar una eficiencia óptima en entornos cotidianos.

3.3.2. Algoritmos de Aprendizaje Profundo Aplicados a OCR

El aprendizaje profundo ha revolucionado el OCR, mejorando notablemente su precisión y eficiencia. Las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) han permitido que el OCR interprete texto en contextos más complejos, con menos errores.

Según Surana et al. (2022), los modelos de aprendizaje profundo aplicados a OCR pueden reconocer texto impreso y manuscrito, incluso en imágenes de baja calidad, lo que aumenta la versatilidad en aplicaciones de lectura asistida. De igual forma, Rahmati et al. (2020) presenta un caso de OCR para textos persas donde redes neuronales especializadas mejoran la precisión, demostrando la capacidad de adaptar estos enfoques a diversos idiomas y estilos de escritura.

En aplicaciones orientadas a la asistencia visual, Ranjan et al. (2021) explica cómo el aprendizaje profundo incrementa la precisión del OCR en tiempo real. Esta habilidad de interpretar texto de origen diverso (impreso, digital, manuscrito) resulta clave para dispositivos de asistencia que deben adaptarse a múltiples fuentes.

Además, Najam y Faizullah (2023) analiza el impacto del aprendizaje profundo en la corrección posterior a la detección del texto, un aspecto crucial para sistemas de lectura asistida. Esta mejora en la precisión y en la corrección de errores garantiza que el usuario reciba información fiable y útil, reduciendo malinterpretaciones y aumentando la usabilidad de la herramienta.

3.3.3. Redes Neuronales Convolucionales (CNN) en Visión Artificial

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) representan una arquitectura de aprendizaje profundo revolucionaria, especializada en el procesamiento de datos con estructura espacial, particularmente imágenes y videos. Su diseño, inspirado en la organización del córtex visual de los mamíferos, implementa capas convolucionales que detectan patrones jerárquicos mediante filtros especializados que recorren la entrada (Beinat, 2022). Estas capas trabajan en conjunto con operaciones de pooling (submuestreo) para reducir la dimensionalidad y preservar características invariantes a traslaciones o rotaciones, permitiendo que el modelo aprenda progresivamente desde patrones básicos hasta estructuras más complejas.

En el contexto de sistemas de asistencia visual, las CNN han demostrado ser fundamentales para diversas tareas críticas como la detección de objetos, el reconocimiento de texto (OCR) y la descripción de escenas para procesar imágenes en tiempo real, identificando textos impresos, rostros y elementos del entorno para traducirlos en descripciones auditivas precisas (Pujante, 2023). Arquitecturas modernas como ResNet o EfficientNet han optimizado el balance entre precisión y eficiencia computacional, aspecto crucial para aplicaciones en dispositivos portátiles con recursos limitados.

La robustez de las CNN frente a variaciones en condiciones reales (iluminación, orientación, escala) las hace particularmente valiosas en escenarios cotidianos. Su capacidad para manejar entornos cambiantes se debe a la naturaleza jerárquica de su procesamiento: las capas iniciales detectan características básicas como bordes y texturas, mientras que las capas más profundas identifican patrones complejos y objetos completos. Esta arquitectura permite filtrar el ruido visual y mejorar la precisión en la detección, aspectos fundamentales cuando el usuario requiere respuestas

inmediatas y confiables.

No obstante, las CNN enfrentan desafíos significativos que requieren consideración. El principal es su necesidad de grandes volúmenes de datos etiquetados para el entrenamiento, junto con su sensibilidad al ruido en las imágenes. Para abordar estas limitaciones, se han desarrollado técnicas como el data augmentation y el transfer learning, que permiten adaptar modelos preentrenados a nuevos contextos, reduciendo significativamente los costos computacionales y mejorando la generalización.

3.3.4. Redes de Memoria a Corto-Largo Plazo (LSTM) en Procesamiento de Secuencias

Las redes de memoria a corto-largo plazo (LSTM) representan una evolución significativa en el campo de las redes neuronales recurrentes (RNN), diseñadas específicamente para manejar secuencias temporales complejas. Su arquitectura distintiva incorpora celdas de memoria y mecanismos de puertas (entrada, olvido y salida) que regulan el flujo de información, permitiendo retener contextos relevantes a largo plazo mientras descartan datos irrelevantes. Esta capacidad las hace especialmente eficaces para superar el problema del gradiente desvaneciente, una limitación común en las RNN tradicionales (Hochreiter & Schmidhuber, 1997).

En el ámbito del procesamiento de texto y sistemas de asistencia, las LSTM demuestran su valor en múltiples aplicaciones. Por ejemplo, en sistemas de OCR, después de que una CNN detecta caracteres en una imagen, la LSTM analiza la secuencia para corregir confusiones entre caracteres similares (como o y 0) o predecir palabras incompletas, considerando el contexto lingüístico completo. Esta capacidad de procesamiento secuencial resulta particularmente valiosa en el reconocimiento de texto manuscrito o en condiciones de baja calidad de imagen (Staudemeyer & Morris, 2019).

La integración de LSTM con CNN ha demostrado ser especialmente efectiva en aplicaciones de asistencia visual. Mientras las CNN extraen características visuales, las LSTM procesan la información secuencial resultante para generar descripciones coherentes y naturales. Esta sinergia mejora significativamente la precisión en tareas como el OCR multilingüe y la generación de

narrativas fluidas para la descripción de escenas, considerando tanto el contexto visual como el lingüístico.

Un desafío significativo en la implementación de LSTM es su demanda computacional, especialmente al procesar secuencias largas. Para abordar esta limitación, se han desarrollado variantes arquitectónicas como las LSTM bidireccionales y los transformadores híbridos, que optimizan el equilibrio entre precisión y velocidad de procesamiento. Además, su integración en sistemas embebidos ha impulsado el desarrollo de técnicas de cuantización y optimización de modelos que mantienen un alto rendimiento mientras reducen los requisitos computacionales.

La combinación de CNN y LSTM en sistemas de asistencia representa un enfoque integral que aprovecha las fortalezas complementarias de ambas arquitecturas. Mientras las CNN sobresalen en la extracción de características espaciales y la comprensión de imágenes, las LSTM destacan en el procesamiento de dependencias temporales y secuenciales. Esta sinergia ha permitido el desarrollo de sistemas más robustos y versátiles, capaces de proporcionar asistencia más precisa y natural a usuarios con discapacidad visual.

4. Herramientas y tecnologías empleadas

4.1. Software

4.1.1. Python

Python es un lenguaje de programación interpretado de alto nivel, diseñado para priorizar la legibilidad del código y la simplicidad de su sintaxis. Creado por Guido van Rossum y lanzado oficialmente en 1991, Python ha evolucionado hasta convertirse en una herramienta esencial en diversas áreas, como el desarrollo de software, el análisis de datos, la inteligencia artificial y la visión por computadora (Foundation, 2021).

La popularidad de Python radica en su ecosistema rico en bibliotecas y *frameworks*, lo que permite a los desarrolladores abordar problemas complejos de manera eficiente. Su sintaxis clara y

accesible reduce la curva de aprendizaje, fomentando la colaboración entre equipos multidisciplinarios. Además, su compatibilidad multiplataforma garantiza que las aplicaciones desarrolladas puedan ejecutarse en una variedad de sistemas operativos, como Windows, macOS y Linux.

Python admite múltiples paradigmas de programación, incluyendo programación orientada a objetos, funcional e imperativa, lo que lo convierte en una herramienta adaptable a diferentes necesidades. Entre las bibliotecas más destacadas se encuentran `OpenCV` para el procesamiento de imágenes, `Tesseract` para el reconocimiento óptico de caracteres, `spaCy` para el procesamiento de lenguaje natural y `gTTS` para la conversión de texto a voz.

Gracias a la robustez de su ecosistema de librerías y a su facilidad de escritura, Python permitió prototipar e integrar todas las funciones de forma ágil en la Raspberry Pi.

4.1.2. Tesseract OCR

Tesseract es una herramienta de reconocimiento óptico de caracteres (OCR) ampliamente utilizada en aplicaciones de procesamiento de texto e imágenes. Sus orígenes se remontan a la década de 1980, cuando fue desarrollada por Hewlett-Packard entre 1985 y 1994. Posteriormente, en 2005, fue liberado como software de código abierto, permitiendo que desarrolladores y académicos adaptaran su tecnología a múltiples necesidades. Desde 2006, Tesseract ha contado con el respaldo de Google, lo cual ha impulsado mejoras notables en su rendimiento y funcionalidad (Smith, 2007).

Una de las características más destacadas de Tesseract es su capacidad para reconocer más de 100 idiomas, lo que la convierte en una solución versátil para entornos multilingües. Además, admite formatos de imagen como PNG, JPEG y TIFF, asegurando una amplia compatibilidad con varios tipos de archivos. En cuanto a la salida, Tesseract ofrece opciones como texto plano, hOCR (HTML), PDF y TSV, facilitando su integración con diferentes sistemas y plataformas.

A lo largo de su evolución, Tesseract ha incorporado mejoras significativas. La versión 4.0, por ejemplo, introdujo un motor basado en redes neuronales LSTM (Long Short-Term Memory), optimizado para el reconocimiento de líneas de texto, al tiempo que conservó el motor tradicional de la versión 3 (basado en patrones de reconocimiento de caracteres). En la actualidad, la versión

5 se considera la más estable y madura (lanzada el 30 de noviembre de 2021), lo que consolida a Tesseract como una de las herramientas OCR más robustas y eficientes disponibles (Smith et al., 2024).

Principios de Funcionamiento de Tesseract. Tesseract sigue un flujo de etapas para convertir una imagen en texto digital:

1. **Preprocesamiento de la Imagen**: Se realiza conversión a escala de grises, binarización (usualmente con el método de Otsu) y eliminación de ruido.
2. **Segmentación**: Divide la imagen en bloques, luego en líneas y finalmente separa palabras y caracteres.
3. **Reconocimiento con LSTM**: Cada carácter (o secuencia de caracteres) se analiza a través de redes neuronales LSTM, que capturan la dependencia entre caracteres para mejorar la exactitud.
4. **Corrección y Post-procesamiento**: Emplea un diccionario interno y reglas gramaticales para minimizar errores, reajustando la segmentación si es necesario.

Configuraciones Clave.

- `-oem` (*Optical Engine Mode*): Selecciona entre el motor de reconocimiento tradicional, el motor LSTM o la combinación de ambos.
- `-psm` (*Page Segmentation Mode*): Ajusta la forma en que Tesseract segmenta la página; por ejemplo, `-psm 6` asume líneas de texto más o menos uniformes.
- `-l` (*Language*): Indica el idioma del texto; en este proyecto, se utiliza `-l spa` para español.

En esta investigación, Tesseract se integra con `pytesseract` para automatizar el reconocimiento de texto una vez que el modelo EAST determina las regiones con mayor probabilidad de contener caracteres.

4.1.3. gTTS (Google Text-to-Speech)

gTTS (Google Text-to-Speech) es una biblioteca de Python diseñada para la conversión de texto a voz (TTS). Esta herramienta hace uso de la API de Google TTS, que se apoya en tecnologías de procesamiento de lenguaje natural (NLP) y aprendizaje profundo para generar voces claras y naturales. Su interfaz sencilla facilita su implementación en diversos proyectos, desde aplicaciones educativas hasta dispositivos de asistencia tecnológica (Karmel et al., 2019).

Entre las ventajas más destacadas de gTTS se incluye su compatibilidad con múltiples idiomas y dialectos, como español, inglés, francés y alemán, entre otros. Esta cualidad resulta especialmente útil en aplicaciones multilingües y en sistemas de asistencia para personas con discapacidad visual, al permitir la conversión de texto impreso o digital en voz sintetizada. Asimismo, gTTS posibilita guardar el audio en formato MP3, lo que garantiza la reproducción en una amplia gama de dispositivos (Sangpal et al., 2019).

4.1.4. Modelo de Detección de Texto EAST

El modelo EAST (*Efficient and Accurate Scene Text Detector*) tiene como objetivo localizar texto en imágenes, incluso en contextos complejos con orientación arbitraria (Zhou et al., 2017). Su salida consiste en cajas delimitadoras (*bounding boxes*) que encapsulan las regiones de texto, pudiendo ser rectángulos rotados o cuadrángulos generales.

Arquitectura y Funcionamiento.

1. **Extracción de Características**: Mediante una red convolucional (usualmente basada en VGG16 o PVANet), la imagen se procesa a múltiples escalas, generando mapas de características ($F1, F2, F3, F4$) que capturan desde detalles locales hasta el contexto global.

2. **Fusión de Características (U-Net style)**: Se combinan las distintas escalas para lograr un mapa unificado que represente tanto textos pequeños como grandes.

3. **Mapas de Puntuación y Geometría**: La capa de salida genera:

- *Mapa de puntuación*: Probabilidad de que cada píxel pertenezca a una región de texto.
- *Mapa de geometría*: Coordenadas de las cajas (rectángulos rotados o cuadrángulos).

4. **Supresión No Máxima (NMS)**: Se refinan los resultados para eliminar superposiciones redundantes, conservando solo las cajas finales de texto.

Función de Pérdida en EAST.

- **Pérdida de Puntuación (L_s)**: Basada en entropía cruzada para evaluar qué tan bien se identifica el texto frente al fondo.
- **Pérdida de Geometría (L_g)**: Mide la diferencia entre las cajas predichas y las reales, empleando *Intersection over Union* (IoU) y métricas L1 normalizadas.

La pérdida total se define como $L = L_s + \lambda_g L_g$, donde λ_g ajusta la importancia relativa de cada término.

EAST se incorpora a través de la API `cv2.dnn` de `OpenCV`, cargando un modelo *preentrenado* en formato `.pb`. Con la finalidad de determinar la región de texto que luego se recorta y envía a Tesseract para el OCR.

4.1.5. OpenCV (Open Source Computer Vision)

OpenCV es una biblioteca de software de código abierto que ha transformado el desarrollo de aplicaciones en el ámbito del procesamiento de imágenes y la visión por computadora. Desde su creación por Intel en 1999, ha ido evolucionando hasta convertirse en una herramienta fundamental para investigadores, desarrolladores e ingenieros que buscan soluciones de visión artificial ágiles y eficientes (Culjak et al., 2012).

Uno de sus hitos más significativos tuvo lugar en la versión 2.0, lanzada en 2009, cuando se adoptó una estructura modular y una interfaz moderna en C++. Este cambio simplificó la organización de sus funcionalidades principales y mejoró la escalabilidad de la biblioteca. Actualmente, OpenCV ofrece más de 2,500 algoritmos optimizados que abarcan desde la detección de objetos y

el análisis de movimiento, hasta la segmentación de imágenes y la reconstrucción tridimensional (Team, 2024).

La flexibilidad de OpenCV se pone de manifiesto en su aplicación en campos tan variados como el análisis de imágenes médicas, la conducción autónoma, la robótica y la seguridad. Asimismo, provee herramientas clave como transformadas de Fourier, operaciones morfológicas y detección de bordes mediante el algoritmo de Canny, fundamentales para diversos proyectos de análisis de imágenes (Praveen & Madihabanu, 2023).

Dado que OpenCV corre de manera eficiente en la Raspberry Pi y brinda funciones extensas de visión, resulta esencial en la lectura asistida propuesta.

4.1.6. spaCy

spaCy es una biblioteca avanzada de procesamiento de lenguaje natural (NLP) desarrollada en Python, concebida para ofrecer eficiencia y facilidad de uso. Creada por la empresa Explosion y lanzada en 2015, spaCy atiende a diversas necesidades de NLP como extracción de información, etiquetado gramatical y reconocimiento de entidades (AI, 2024).

Entre sus principales características, spaCy incluye modelos estadísticos preentrenados para varios idiomas y una arquitectura extensible mediante *plugins* y componentes adicionales. JUGRAN et al. (2021) demostraron la versatilidad de spaCy al usarlo en sistemas de resumen automático de texto, evidenciando su eficacia en distintas tareas de análisis.

Las capacidades técnicas de spaCy se basan en el uso de redes neuronales convolucionales y algoritmos de *parsing* incremental, lo que se traduce en un alto nivel de eficiencia y precisión (AI, 2024). En proyectos de accesibilidad, la librería puede emplearse para la corrección semántica o la simplificación del texto reconocido, mejorando la experiencia de lectura al convertir el contenido a formato de audio.

Funciones principales.

1. Analiza la secuencia de palabras (*tokens*), corrigiendo uniones o separaciones erróneas.

2. Aplica sus capacidades de *parsing* para eliminar símbolos extraños o ruidos no deseados.

3. Asegura una mayor coherencia léxica.

4.1.7. pynput

`pynput` es una biblioteca de Python diseñada para la captura y el control de dispositivos de entrada, como teclados y ratones. Esta herramienta, desarrollada con un enfoque multiplataforma, permite monitorear eventos de entrada en tiempo real, lo que la hace ideal para aplicaciones que requieren interacción dinámica entre el usuario y el sistema (Palmér, 2024).

Entre sus principales funcionalidades, `pynput` permite capturar eventos de teclas presionadas, teclas liberadas y combinaciones específicas. También proporciona una API intuitiva para simular acciones de teclado o ratón, lo que la convierte en una herramienta versátil para la automatización de tareas y el desarrollo de interfaces interactivas.

La biblioteca es compatible con sistemas operativos como Windows, macOS y Linux, y ofrece una integración sencilla con otros módulos de Python. Gracias a su documentación detallada y su diseño accesible, `pynput` se ha convertido en una opción popular entre desarrolladores que buscan implementar soluciones de interacción basadas en dispositivos de entrada.

4.1.8. VLC Media Player

VLC Media Player es un reproductor multimedia de código abierto desarrollado por el proyecto VideoLAN. Desde su lanzamiento inicial en 2001, VLC ha destacado por su versatilidad y capacidad para reproducir una amplia variedad de formatos de audio y video sin necesidad de instalar códecs adicionales (VideoLAN, 2024). Su diseño modular y su soporte para múltiples plataformas, como Windows, macOS, Linux y sistemas móviles, lo convierten en una herramienta confiable en el manejo de contenido multimedia.

Una de las características más destacadas de VLC es su capacidad para integrarse con otros sistemas mediante interfaces programáticas. Esto es posible gracias a bibliotecas como `python-vlc`, que permite el control de funciones como reproducción, pausa, reanudación y ajuste de volumen me-

diante scripts de Python. Además, VLC admite una amplia gama de formatos de archivo, incluidos MP3, AAC, FLAC, MP4 y AVI, garantizando su compatibilidad con diversos tipos de medios.

Por su carácter de código abierto y su licencia GNU GPL, VLC es ampliamente utilizado tanto en aplicaciones comerciales como académicas. Su arquitectura flexible y su fiabilidad lo convierten en una de las herramientas multimedia más populares y robustas en la actualidad.

4.2. Hardware

4.2.1. Raspberry Pi 4 (Modelo B, 4 GB RAM)

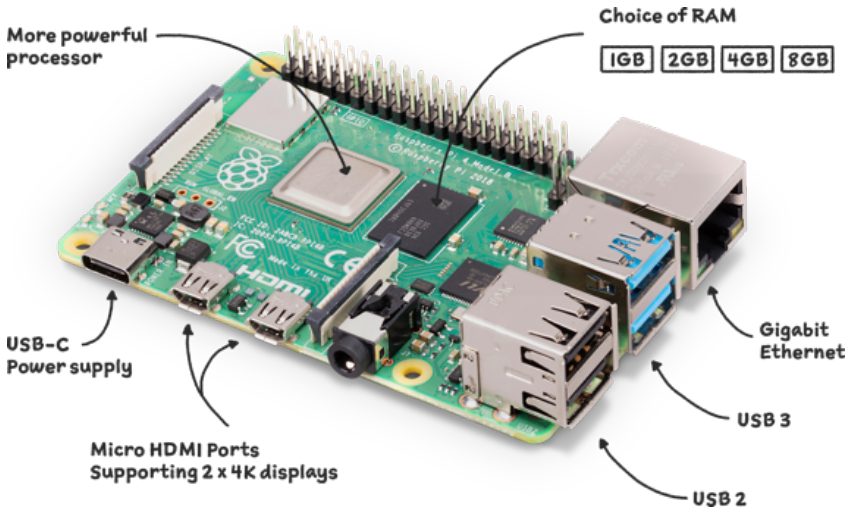
La Raspberry Pi 4, ilustrada en la Figura 5, es una computadora de placa única diseñada para ofrecer un alto rendimiento en un formato compacto. Equipada con un procesador Broadcom BCM2711 de cuatro núcleos Cortex-A72 a 1.5 GHz y 4 GB de RAM, esta placa es ideal para ejecutar aplicaciones que requieren procesamiento intensivo, como visión por computadora y reconocimiento óptico de caracteres (Raspberry Pi Foundation, 2019).

Especificaciones técnicas:

- **Procesador:** Broadcom BCM2711, ARM Cortex-A72 (64 bits) de cuatro núcleos a 1.5 GHz.
- **Memoria RAM:** 4 GB LPDDR4.
- **Conectividad:**
  - 2 puertos USB 3.0 y 2 puertos USB 2.0.
  - Conector Ethernet Gigabit.
  - Bluetooth 5.0 y Wi-Fi de doble banda (2.4 GHz y 5 GHz).
- **Alimentación:** Puerto USB-C (5 V, 3 A).
- **Salidas de video y audio:** 2 puertos micro-HDMI con soporte para resolución 4K, y un conector de audio analógico de 3.5 mm.
- **Dimensiones:** 85.6 x 56.5 mm.

Figura 5

Raspberry Pi 4 (Modelo B, 4 GB RAM).



Nota: Imagen obtenida de la página oficial de Raspberry Pi (Raspberry Pi Foundation, 2019).

4.2.2. Cámara Logitech C920s

La cámara Logitech C920s, expuesta en la Figura 6, es conocida por su calidad de imagen y fiabilidad, lo que la convierte en una opción ideal para tareas de captura de imágenes en alta resolución. Este modelo ofrece enfoque automático y un campo de visión de 78°, garantizando imágenes nítidas incluso en condiciones de iluminación variables (Logitech, 2019).

Especificaciones técnicas:

- **Resolución máxima:** 1080p a 30 fps.
- **Enfoque:** Automático.
- **Conectividad:** USB 2.0.
- **Campo de visión (FOV):** 78° (aproximadamente).

Gracias a su nitidez y capacidad para capturar detalles precisos, la Logitech C920s es ampliamente utilizada en aplicaciones de transmisión, videoconferencias y análisis de imágenes.

Figura 6

Cámara Logitech C920s.

Nota: Imagen obtenida de Logitech (2019).

4.2.3. Parlantes Genius SP-U115

Los parlantes Genius SP-U115, mostrados en la Figura 7, son una solución compacta y eficiente para la salida de audio. Alimentados mediante USB, estos altavoces ofrecen un nivel de volumen adecuado para la reproducción de contenido multimedia, siendo ideales para sistemas portátiles y embebidos (Genius, 2020).

Especificaciones técnicas:

- **Potencia de salida:** 3 W RMS (aproximadamente).
- **Alimentación:** USB.
- **Conexión de audio:** Jack de 3.5 mm.

Su diseño compacto y facilidad de conexión los convierten en una opción conveniente para aplicaciones que requieren reproducción de audio clara y consistente.

Figura 7

Parlantes Genius SP-U115.

Nota: Imagen adaptada de Genius (2020).

4.2.4. Teclado Numérico Super Slim Numeric Keypad

El teclado Super Slim Numeric Keypad es un dispositivo compacto diseñado para facilitar la entrada de datos y comandos en sistemas embebidos. Su diseño ergonómico y su compatibilidad multiplataforma lo hacen ideal para tareas específicas que requieren una interacción rápida y precisa.

Especificaciones técnicas:

- **Alimentación:** USB (5 V).
- **Compatibilidad:** Windows, Linux, macOS.
- **Dimensiones:** Compacto y ligero, ideal para entornos de espacio reducido.

Este teclado numérico simplifica la interacción con sistemas que requieren entradas específicas, siendo especialmente útil en aplicaciones de automatización y accesibilidad.

Figura 8

Teclado Numérico Super Slim.

Nota: Teclado numérico compacto y versátil, diseñado para interacción en sistemas embebidos extraído de Techly (2023).

5. Metodología

5.1. Diseño del Sistema

La Figura 12 presenta una visión integral de cómo interactúan los diferentes componentes del sistema para lograr la funcionalidad deseada. Este diagrama destaca las interacciones desde la activación del sistema hasta la reproducción de audio, pasando por la captura y procesamiento de imágenes, así como la detección y reconocimiento de texto.

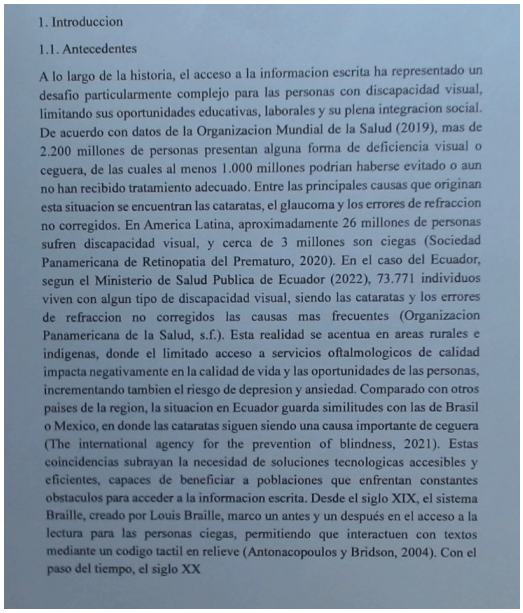
Proceso de Implementación:

1. **Inicio: Activación del Sistema (Raspberry Pi):** El sistema comienza a operar automáticamente al encender la Raspberry Pi, ejecutando un script principal que coordina todas las tareas. Este script inicializa los módulos necesarios, como la cámara, los modelos de detección de texto y las librerías de conversión de texto a voz. Además, se reproduce un mensaje de bienvenida pregrabado para informar al usuario que el sistema está listo para su uso.

2. **Captura de Imagen (Cámara Logitech C920s):** La Figura 9 demuestra la imagen capturada utilizando la cámara Logitech C920s, se implementa una función en Python junto con `OpenCV` que activa la cámara y captura el frame en respuesta a un evento específico.

Figura 9

Captura de una imagen utilizando la cámara del dispositivo.


1. Introduccion

1.1. Antecedentes

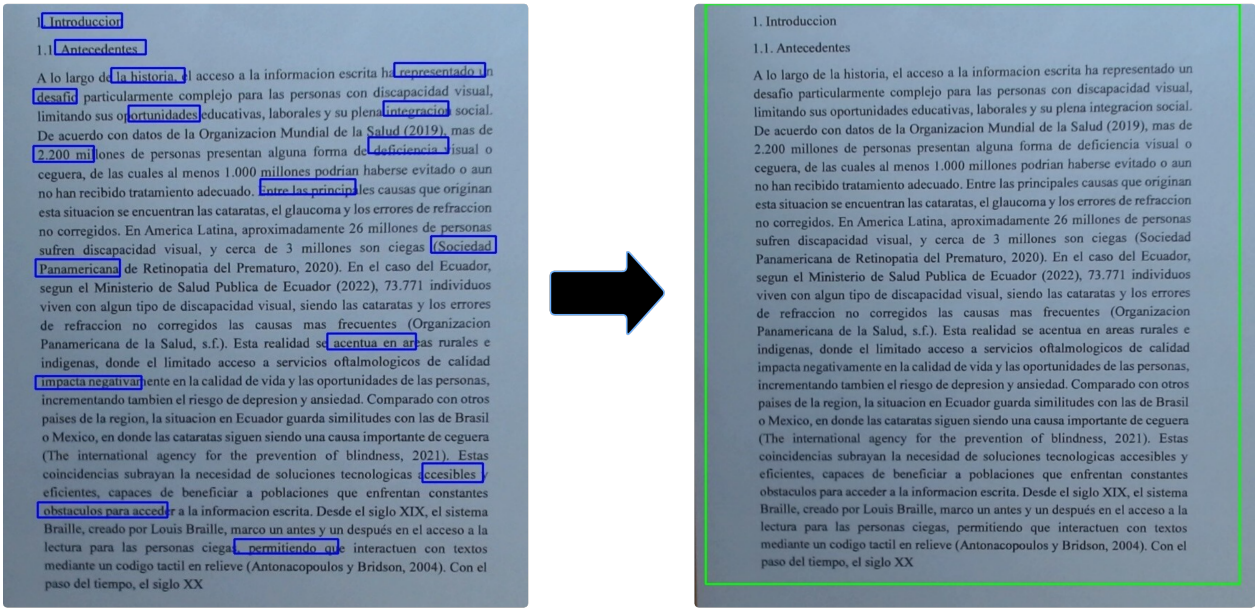
A lo largo de la historia, el acceso a la informacion escrita ha representado un desafio particularmente complejo para las personas con discapacidad visual, limitando sus oportunidades educativas, laborales y su plena integracion social. De acuerdo con datos de la Organizacion Mundial de la Salud (2019), mas de 2.200 millones de personas presentan alguna forma de deficiencia visual o ceguera, de las cuales al menos 1.000 millones podrian haberse evitado o aun no han recibido tratamiento adecuado. Entre las principales causas que originan esta situacion se encuentran las cataratas, el glaucoma y los errores de refraccion no corregidos. En America Latina, aproximadamente 26 millones de personas sufren discapacidad visual, y cerca de 3 millones son ciegas (Sociedad Panamericana de Retinopatia del Prematuro, 2020). En el caso del Ecuador, segun el Ministerio de Salud Publica de Ecuador (2022), 73.771 individuos viven con algun tipo de discapacidad visual, siendo las cataratas y los errores de refraccion no corregidos las causas mas frecuentes (Organizacion Panamericana de la Salud, s.f.). Esta realidad se acentua en areas rurales e indigenas, donde el limitado acceso a servicios oftalmologicos de calidad impacta negativamente en la calidad de vida y las oportunidades de las personas, incrementando tambien el riesgo de depresion y ansiedad. Comparado con otros paises de la region, la situacion en Ecuador guarda similitudes con las de Brasil o Mexico, en donde las cataratas siguen siendo una causa importante de ceguera (The international agency for the prevention of blindness, 2021). Estas coincidencias subrayan la necesidad de soluciones tecnologicas accesibles y eficientes, capaces de beneficiar a poblaciones que enfrentan constantes obstaculos para acceder a la informacion escrita. Desde el siglo XIX, el sistema Braille, creado por Louis Braille, marco un antes y un después en el acceso a la lectura para las personas ciegas, permitiendo que interactuen con textos mediante un codigo tactil en relieve (Antonacopoulos y Bridson, 2004). Con el paso del tiempo, el siglo XX

Nota: La imagen muestra la vista previa capturada por la cámara antes de procesar el texto.

3. **Detección de Texto (Modelo EAST):** Como se observa en la Figura 10, una vez capturada la imagen, se utiliza el modelo EAST (Efficient and Accurate Scene Text Detector) para detectar regiones de texto en la imagen. Este proceso implica redimensionar la imagen, generar un blob para el modelo y realizar la inferencia para obtener las áreas donde se encuentra texto.

Figura 10

La cámara del dispositivo detectando texto, resaltando los caracteres mediante un recuadro delimitador (verde) para indicar el área de reconocimiento exitoso el cual abarca todo el texto de manera correcta.

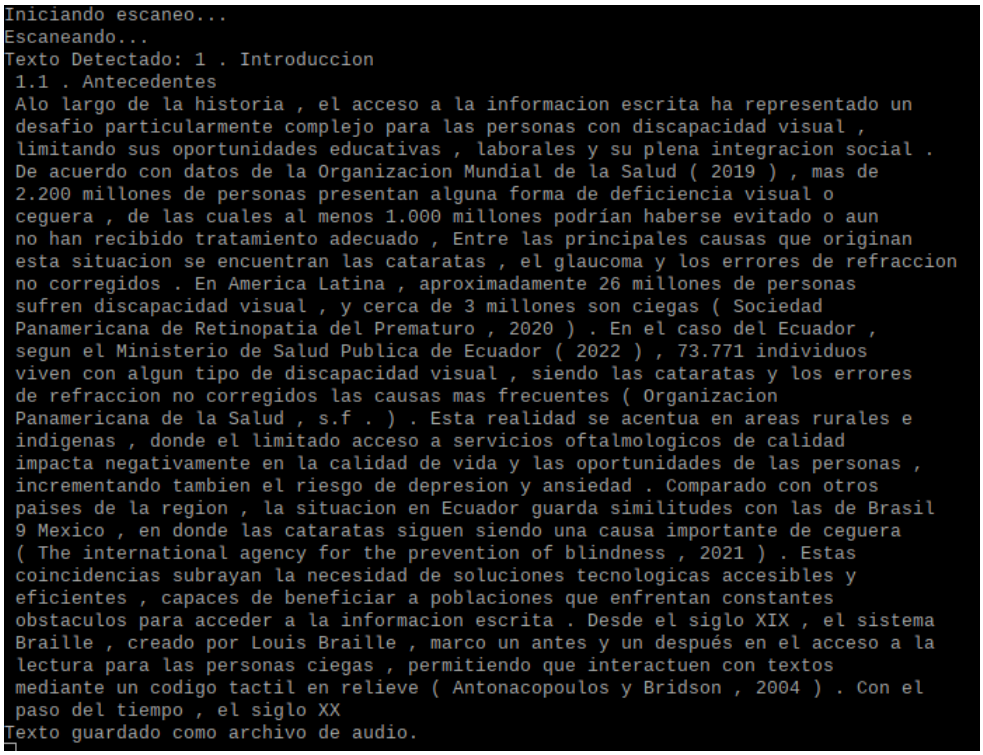


En esta imagen, la cámara integrada en el dispositivo está en funcionamiento, capturando el texto de una hoja impresa.

4. **Verificación de Detección de Texto:** Se verifica si se detectó texto en la imagen.
    - **Sí:** Si se detecta texto, se procede al reconocimiento del mismo utilizando Tesseract OCR.
    - **No:** Si no se detecta texto, el sistema vuelve a hacer una captura de imagen.
5. **Reconocimiento de Texto (Tesseract OCR):** Con las coordenadas de las región de texto detectada, Tesseract OCR, en la Figura 11 procesa la región global para extraer el texto digitalmente.

Figura 11

Resultado del reconocimiento óptico de caracteres (OCR) utilizando Tesseract.



Nota: La imagen muestra el texto reconocido por el sistema OCR, que se imprime directamente en la terminal de la Raspberry Pi.

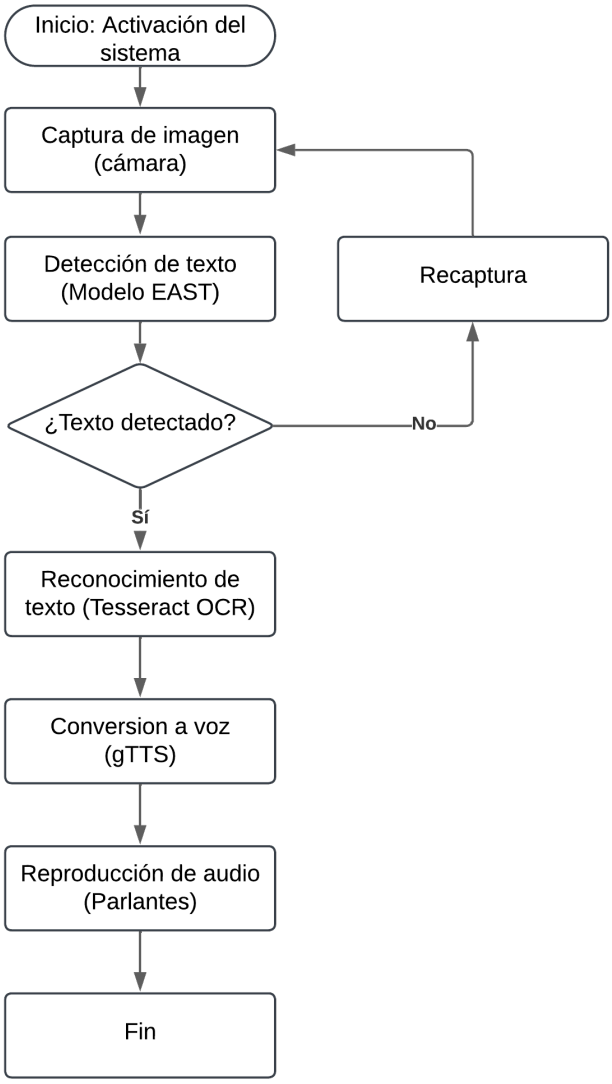
6. **Conversión a Voz (gTTS):** El texto reconocido se pasa a la herramienta gTTS (Google Text-to-Speech) para convertirlo en un archivo de audio.

7. **Reproducción de Audio (Parlantes Genius SP-U115):** El archivo de audio generado se reproduce a través de los parlantes Genius SP-U115, permitiendo al usuario escuchar el contenido textual.

8. **Fin:** Tras la reproducción el sistema vuelve a esperar a la próxima orden del usuario

Figura 12

Diagrama de flujo general del sistema.



Nota: Este diagrama ilustra el flujo completo desde la activación del sistema hasta la reproducción de audio, incluyendo las decisiones clave que determinan el siguiente paso en el proceso.

5.1.1. Implementación de los Módulos

A continuación, se describen los módulos más relevantes: captura de imagen, detección de texto con EAST, reconocimiento de texto con Tesseract, conversión de texto a voz con gTTS,

reproducción de audio con VLC, y el listener de teclado con `pynput`.

5.1.2. Captura de Imagen

Este módulo asegura que las imágenes sean de alta calidad, adecuadas para la posterior detección de texto. Para la captura de imágenes, se utilizó los siguientes parámetros de configuración:

- **Resolución de Captura:** 1920x1080píxeles (HD).
- **Frames por Segundo (FPS):** 30 FPS.
- **Almacenamiento Temporal:** Las imágenes capturadas se almacenan temporalmente en una carpeta designada antes de ser procesadas.

5.1.3. Detección de Texto con Modelo EAST

La Figura 13 muestra el proceso del modelo EAST el cual se implementó utilizando la librería `OpenCV`, al ser un modelo preentrenado, no se requirió una fase de entrenamiento manual, el cual esta cargado en formato `.pb` mediante su módulo `cv2.dnn`. El proceso comienza con la redimensión de la imagen capturada a 640x640 píxeles, optimizando así el rendimiento del modelo EAST. Tras generar el blob necesario, se realiza la inferencia para detectar las regiones de texto en la imagen. La aplicación de Non-Max Suppression (NMS) permite refinar las detecciones, eliminando redundancias y asegurando precisión en la identificación de texto.

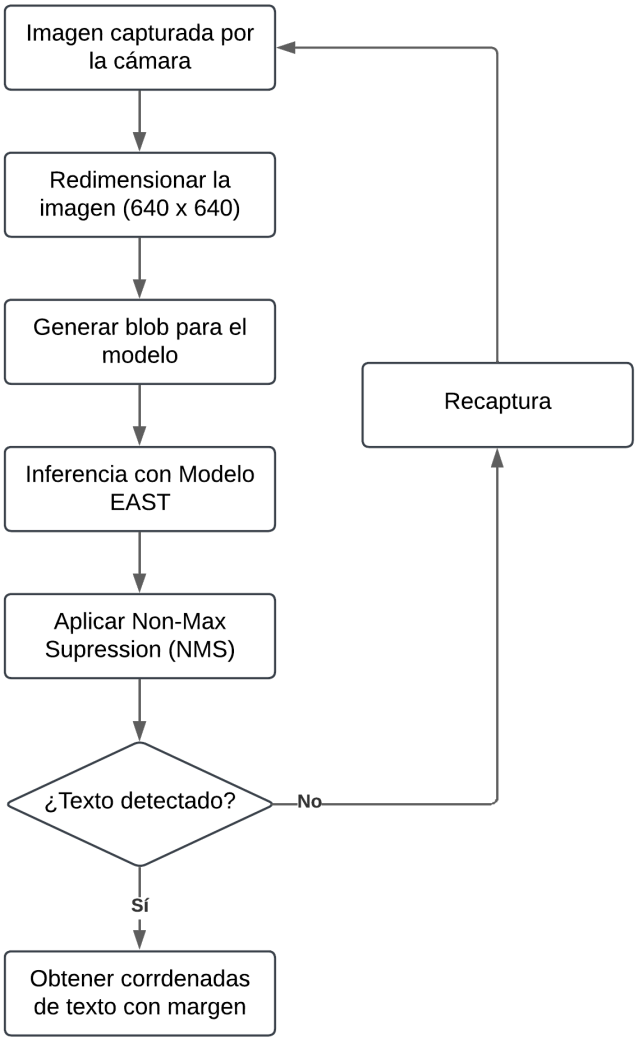
Parámetros de configuración:

- **Threshold de Confianza:** 0.5, utilizado para filtrar detecciones con baja probabilidad de contener texto.
- **Generación de Cuadro Delimitador Global:** Se crea un cuadro delimitador global alrededor de todas las regiones de texto detectadas, agregando un margen adicional del 10 % para asegurar que se abarque todo el texto de manera efectiva.

- **Inferencia del Modelo:** Utilización de `cv2.dnn.forward()` para obtener los mapas de puntuación y geometría que identifican las regiones de texto.

Figura 13

Flujo detallado del modelo EAST.



Nota: Representa el proceso de detección de texto, incluyendo la redimensión de la imagen, generación de blobs, inferencia del modelo, y la aplicación de Non-Max Suppression (NMS).

5.1.4. Reconocimiento de Texto con Tesseract OCR

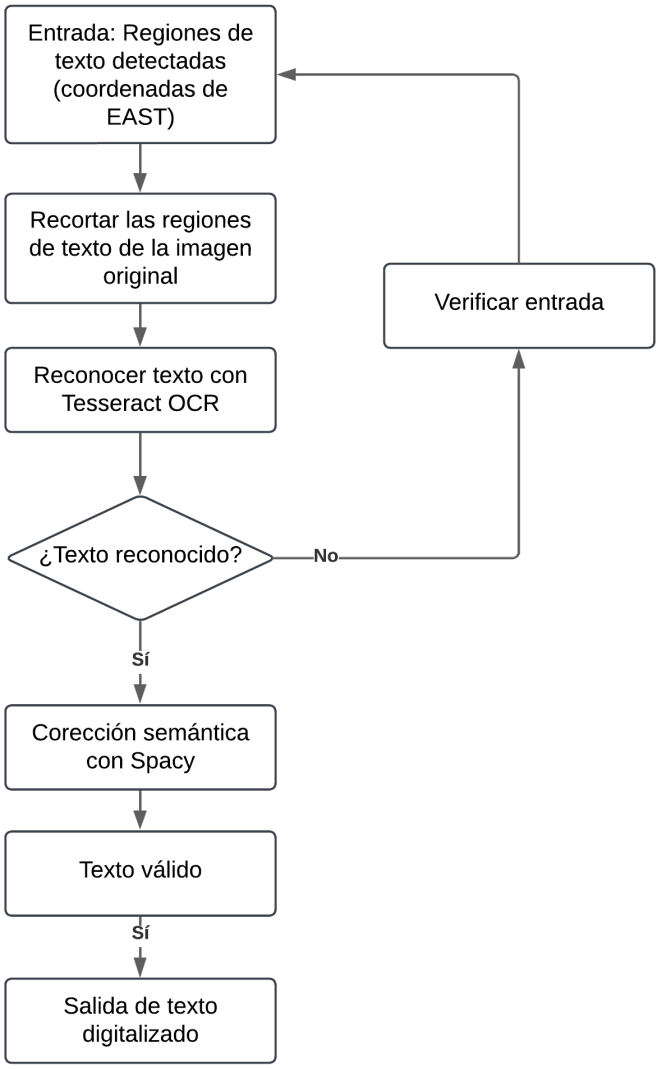
La Figura 14 ilustra el proceso en el cual, una vez detectadas las regiones de texto, estas se recortan de la imagen original y se procesan mediante `Tesseract OCR`, empleando la librería `pytesseract`, fundamentada en redes neuronales LSTM, que extrae el contenido textual carácter por carácter. para integrar esta herramienta en el flujo de trabajo de Python. El texto resultante del OCR puede presentar saltos de línea no deseados o ligeros errores de segmentación. Para refinar dicha salida, se aplicó spaCy en un paso posterior. Este módulo de Procesamiento de Lenguaje Natural revisa la secuencia reconocida, normaliza tokenizaciones y corrige la unión o separación inexacta de palabras, especialmente útil en textos con guiones o faltas de ortografía mínimas. Aunque la corrección no es exhaustiva en términos ortográficos, spaCy contribuye notablemente a la coherencia final.

Parámetros de configuración:

- **Preprocesamiento de Imagen:**
  - **Conversión a Escala de Grises:** Mejora la calidad del reconocimiento al eliminar información de color innecesaria.
- **Configuración de Tesseract:**
  - **Idioma:** Configuración para reconocer texto en español (`lang='spa'`).
  - **Modo de Página:** Modo de bloque de texto (`-psm 6`) para optimizar el reconocimiento de texto en regiones específicas, se limita a un conjunto de cadenas de texto con un mínimo de falsos positivos.

Figura 14

Flujo del proceso en Tesseract OCR.



Nota: Muestra el proceso de reconocimiento de texto, incluyendo la recorte de las regiones de texto, reconocimiento con Tesseract OCR, corrección semántica con SpaCy, y la validación del texto resultante.

5.1.5. Generación de Audios Pregrabados

Para optimizar el flujo del sistema, se generaron audios pregrabados mediante `gTTS` (Google Text-to-Speech). Estos audios incluyen mensajes iniciales y de confirmación, almacenándose como

archivos MP3 en la Raspberry Pi para su uso repetido.

Parámetros de configuración::

- **Mensajes Incluidos:**
  - Mensaje de inicio del sistema.
  - Confirmación de escaneo exitoso.
  - Advertencia de falta de texto detectado.
- **Formato de Audio:** MP3.
- **Idioma:** Español (`lang='es'`).

5.1.6. Conversión de Texto a Voz con gTTS

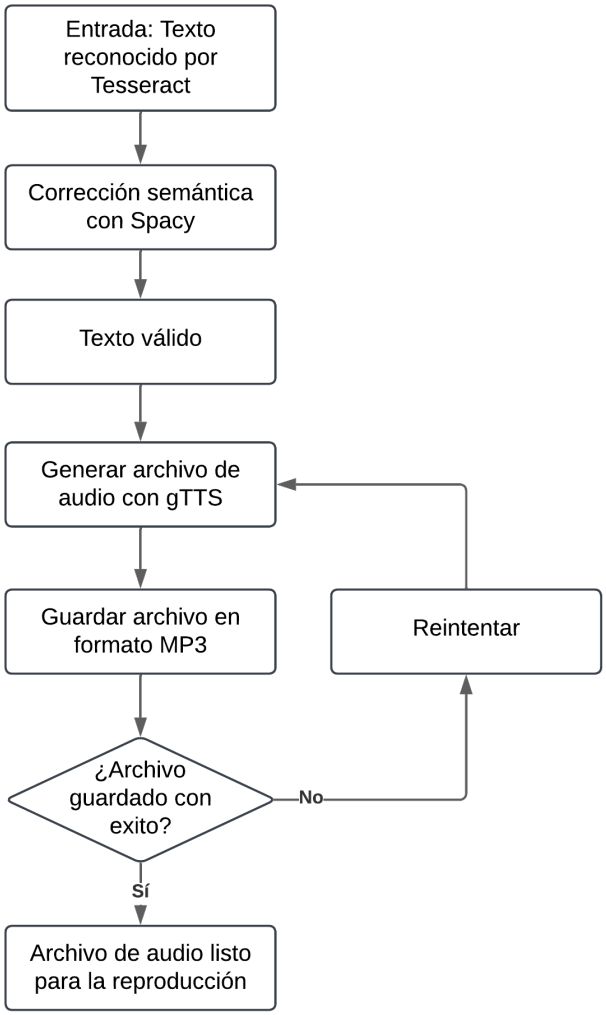
El texto digitalizado y corregido se convierte a voz utilizando `gTTS`. Este módulo genera archivos de audio en formato MP3 que son almacenados en el sistema. Se implementaron verificaciones para asegurar que el texto no esté vacío antes de la conversión y que el archivo de audio se guarde correctamente.

Parámetros de configuración::

- **Configuración de gTTS:**
  - **Idioma:** Español (`lang='es'`).
  - **Velocidad de Voz:** Parámetro de velocidad ajustado para una reproducción natural (`slow=False`).

Figura 15

Flujo del proceso en gTTS.



Nota: Ilustra el proceso de conversión de texto a voz, incluyendo la validación del texto, generación del audio con gTTS, y la gestión del almacenamiento del archivo de audio en formato MP3.

5.1.7. Reproducción de Audio

Para la reproducción de los archivos de audio generados, se utilizó la librería `vlc`, que permite manejar reproducción, pausa y reinicio mediante el objeto `MediaPlayer`. Esto garantiza una experiencia fluida y personalizable para el usuario.

Parámetros de configuración:

- **Inicialización de VLC:** Utilización de `MediaPlayer` para cargar y gestionar los archivos MP3.
- **Control de Reproducción:**
  - **Pausar/Reanudar:** Implementación de funciones para pausar y reanudar la reproducción.
  - **Reiniciar:** Función para detener y reiniciar la reproducción desde el inicio del archivo de audio.

5.1.8. Interacción del Usuario con Listener de Teclado

El listener de teclado fue implementado utilizando la librería `pynput`, permitiendo al usuario interactuar con el sistema mediante la presión de teclas específicas.

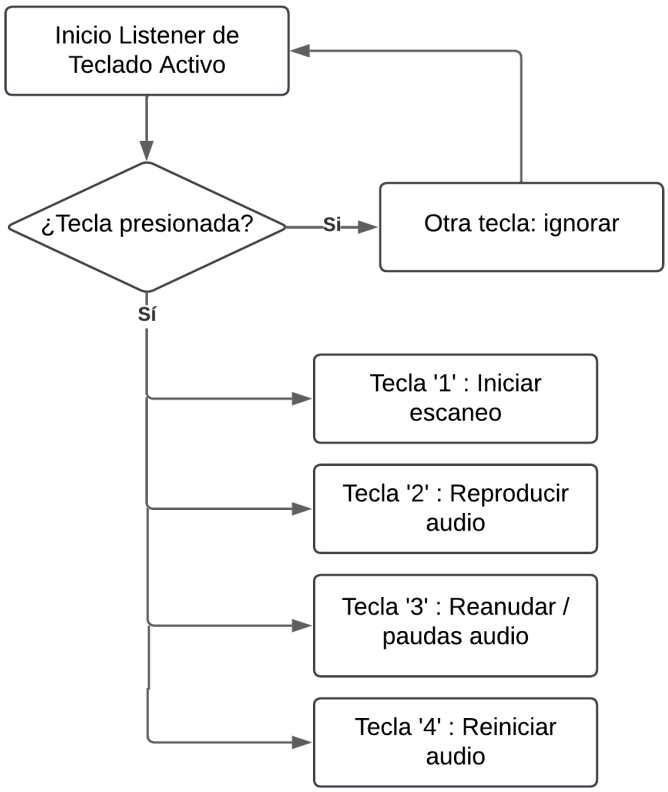
Parámetros de configuración::

- **Mapeo de Teclas:**
  - **Tecla '1':** Iniciar Escaneo.
  - **Tecla '2':** Reproducir Audio.
  - **Tecla '3':** Pausar/Reanudar Audio.
  - **Tecla '4':** Reiniciar Audio.
  - **Tecla '5':** Salir del Sistema.
- **Debounce de Teclas:** Implementación de un mecanismo de debounce para evitar múltiples activaciones no deseadas al mantener presionada una tecla.
- **Funciones Asociadas:** Cada tecla asignada ejecuta una función definida que interactúa con los módulos correspondientes, asegurando una integración fluida entre la entrada del usuario y las acciones del sistema.

- **Gestión de Estados del Sistema:** Se mantiene un estado interno para gestionar correctamente las transiciones entre las diferentes acciones del sistema, evitando conflictos y asegurando una operación coherente.
- **Manejo de Errores:** Implementación de bloques `try-except` para capturar y gestionar excepciones durante la detección de teclas y ejecución de funciones asociadas.

Figura 16

Flujo de interacción del Listener de Teclado.



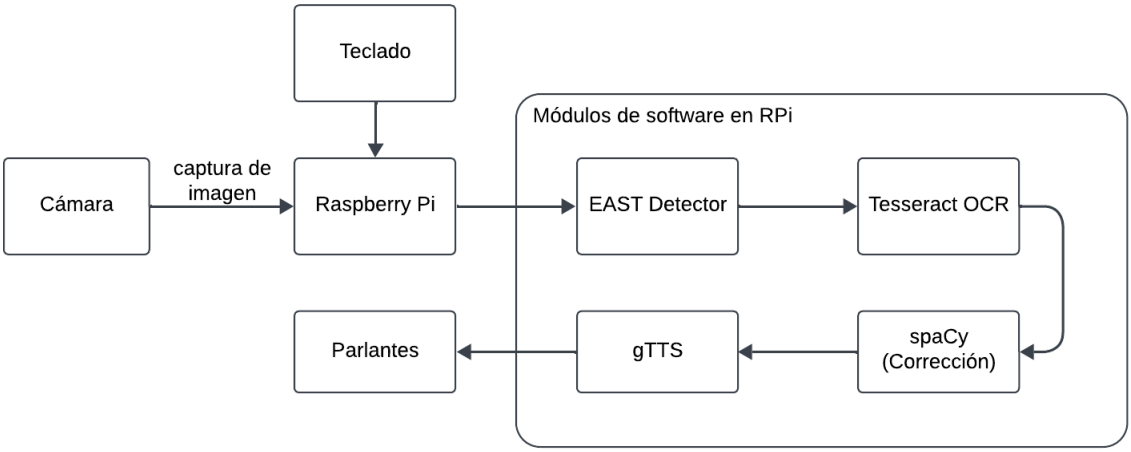
Nota: Representa cómo el listener de teclado detecta las teclas presionadas y ejecuta las acciones correspondientes, asegurando una interacción fluida y eficiente con el usuario.

5.1.9. Integración de Módulos

El flujo completo del sistema está representado en la Figura 17, donde se observan las interacciones entre los módulos de captura de imagen, detección de texto, reconocimiento OCR, generación de audio y reproducción de audio, todo coordinado mediante el listener de teclado para una interacción eficiente con el usuario.

Figura 17

Integración de los Módulos del Sistema.



Nota: Este diagrama ilustra cómo los diferentes módulos del sistema interactúan entre sí, desde la captura de imágenes hasta la reproducción de audio, pasando por la detección y reconocimiento de texto.

Todos los módulos descritos anteriormente están integrados en un script principal que coordina las interacciones entre ellos. Este script gestiona el flujo de trabajo completo, desde la captura de imagen hasta la reproducción de audio, respondiendo a las entradas del usuario a través del listener de teclado. La modularidad del diseño permite una fácil extensión y mantenimiento del sistema, asegurando que cada componente funcione de manera independiente pero coordinada.

6. RESULTADOS Y DISCUSIONES

6.1. Desempeño del OCR bajo Diferentes Condiciones de Iluminación

Para evaluar el desempeño del sistema de Reconocimiento Óptico de Caracteres (OCR) en distintos entornos, se realizaron pruebas regulares bajo tres escenarios de iluminación controlada: iluminación intensa con 1305 lux, iluminación moderada con 222 lux e iluminación reducida con 10 lux, con el objetivo de analizar su eficacia en condiciones variables.

A su vez, se emplearon dos dimensiones tipográficas distintas (12 pt y 14 pt), aplicadas a 10 documentos de prueba en cada condición, con el fin de determinar la respuesta del sistema ante tamaños de fuente y contrastes variables.

Cálculo de Precisión del OCR. La precisión se define como el porcentaje de palabras correctamente reconocidas respecto al total de palabras reales en el documento. Para cada prueba:

$$\text{Precisión } (\%) \;=\; \left(\frac{\text{Número de Palabras Correctas}}{\text{Número Total de Palabras}}\right) \times 100.$$

De esta manera, se generan valores puntuales en un rango de 0–100 %, y posteriormente se calcula el promedio de cada escenario (fuente e iluminación).

Métrica de Claridad de Audio. Adicionalmente, para determinar la comprensibilidad y calidad perceptual de la síntesis de voz generada por gTTS, se utilizó una escala subjetiva de 1 a 5, siendo 1 una calidad muy difícil de entender y 5 un audio fluido y comprensible.

Medición del Tiempo de Procesamiento. El tiempo de procesamiento se midió desde el instante en que se inicia la captura de la imagen hasta que el sistema completa la síntesis del audio. Se registró el tiempo total en segundos, permitiendo compararlo entre distintas condiciones lumínicas y tamaños de fuente. Este parámetro busca reflejar la usabilidad del sistema en la práctica, pues un tiempo excesivo podría limitar su aplicabilidad real en tareas de lectura asistida.

6.2. Iluminación Intensa (1305 lux)

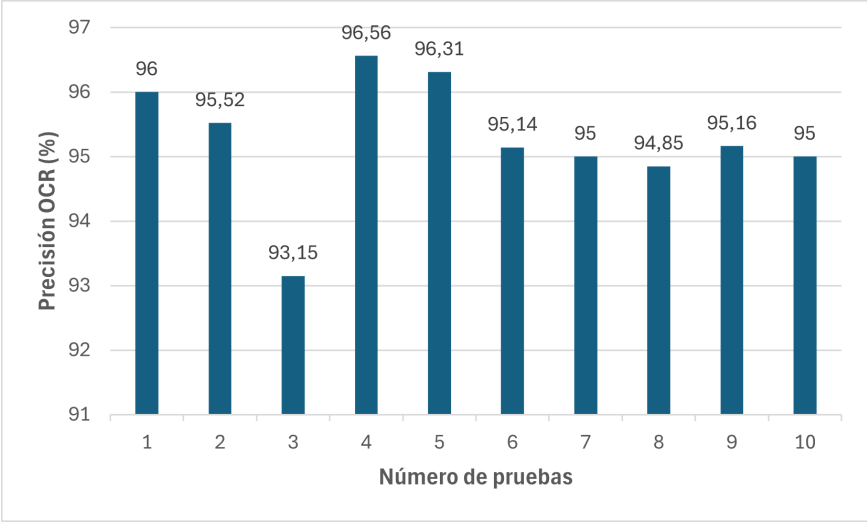
6.2.1. Precisión OCR para Fuente 12 pt

La Figura 18 presenta los resultados del análisis de precisión OCR para la fuente de 12 pt bajo iluminación intensa. Los datos reflejan un desempeño consistente del sistema, con un promedio de precisión del 95,27 %. Esto destaca la capacidad del OCR para reconocer caracteres pequeños de forma precisa en condiciones de iluminación controladas, como oficinas bien iluminadas o laboratorios con luz adecuada.

Aunque el rendimiento es muy similar al de la fuente de 14 pt, se observa que la dimensión menor de los caracteres puede aumentar ligeramente la sensibilidad a errores en situaciones de baja resolución o pequeñas distorsiones generadas por el desenfoque o sombras. Sin embargo, la precisión general lograda en estas pruebas confirma que la fuente de 12 pt es adecuada para aplicaciones donde los textos sean claros y la iluminación sea controlada.

Figura 18

Distribución de precisión OCR para fuente de 12 puntos bajo iluminación intensa.



El análisis contempla 10 iteraciones de prueba bajo condiciones de iluminación intensa (1305 lux), mostrando la variabilidad y consistencia del sistema.

La Tabla 1 presenta los datos numéricos de las diez pruebas realizadas bajo condiciones de iluminación alta (1305 lux), utilizando texto con fuente de 12 pt.

En este escenario, el rango de precisión se ubica aproximadamente entre 93 % (prueba 3) y 96.5 % (prueba 4). Esto implica que la mayoría de las palabras fueron reconocidas correctamente, aun cuando los caracteres son relativamente pequeños (12 pt). El tiempo de procesamiento oscila entre 40 y 46 segundos, un margen aceptable para documentos de 298 a 369 palabras. Asimismo, la claridad del audio es consistentemente alta (de 4.5 a 4.8), confirmando que los errores OCR no afectan drásticamente la pronunciación final.

Tabla 1

Resultados detallados de OCR (Tamaño de Fuente 12 pt) bajo Iluminación Alta (1305 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	369	356	96.00	46	4.5
2	335	321	95.52	44	4.7
3	336	313	93.15	41	4.7
4	320	309	96.56	40	4.6
5	298	286	96.31	45	4.8
6	350	333	95.14	43	4.65
7	340	323	95.00	44	4.7
8	330	313	94.85	42	4.6
9	310	295	95.16	45	4.75
10	300	285	95.00	44	4.7

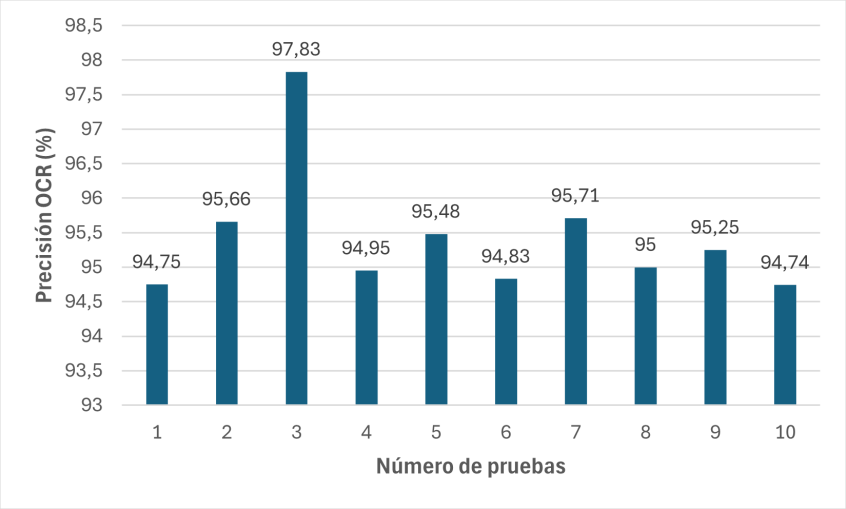
Nota: Se confirma un promedio de 95.27 % de precisión, 43.4 s de tiempo y 4.67 de claridad (escala 1–5).

6.2.2. Precisión OCR para Fuente 14 pt

La Figura 19 demuestra que la implementación del OCR con fuente de 14 pt mantiene un rendimiento sobresaliente, con un promedio de precisión del 95,42 %. Este desempeño refleja la capacidad del sistema para procesar caracteres de mayor tamaño de manera eficiente, contribuyendo a una mayor claridad y facilidad de segmentación durante el análisis de las imágenes. Comparado con la fuente de 12 pt, esta configuración presenta una ventaja leve en términos de precisión, lo que evidencia que los caracteres más grandes son reconocidos con mayor éxito bajo condiciones óptimas de iluminación. Además, la consistencia observada en las pruebas resalta la robustez del sistema frente a pequeñas imperfecciones en los textos, como sombras accidentales o irregularidades menores en las fuentes. Esto refuerza su utilidad para aplicaciones en entornos controlados, como laboratorios o espacios con condiciones de luz constantes, donde la iluminación ideal es un factor primordial. El rendimiento de esta configuración confirma la importancia de adaptar el tamaño de la fuente al contexto de uso y a las condiciones de iluminación.

Figura 19

Distribución de precisión OCR para fuente de 14 puntos bajo iluminación intensa.



Resultados de las iteraciones de prueba bajo condiciones de iluminación intensa (1305 lux), evidenciando el comportamiento del sistema con tipografía de mayor dimensión.

En la Tabla 2 se muestran los valores de precisión, tiempo y claridad para textos de 286 a 300 palabras, en 10 iteraciones realizadas bajo 1305 lux. El tamaño de fuente 14 pt ofrece una ligera ventaja, con picos de precisión cercanos al 97.8 % (prueba 3).

La precisión mínima (94.74 %) sigue siendo superior al 94 %, lo que indica un reconocimiento muy confiable. El tiempo de procesamiento varía entre 41 y 45 segundos, similar a la fuente de 12 pt, evidenciando que el aumento en el tamaño de los caracteres no ralentiza el flujo de OCR. En cuanto a la claridad, los valores rondan 4.8, evidenciando una lectura sintetizada limpia, con menos palabras confusas producto de errores OCR.

Tabla 2

Resultados detallados de OCR (Tamaño de Fuente 14 pt) bajo Iluminación Alta (1305 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	286	271	94.75	45	4.8
2	277	265	95.66	44	4.8
3	278	272	97.83	43	4.9
4	297	282	94.95	42	4.8
5	283	271	95.48	41	4.8
6	290	275	94.83	43	4.8
7	280	268	95.71	44	4.8
8	300	285	95.00	42	4.8
9	295	281	95.25	43	4.8
10	285	270	94.74	44	4.8

Nota: El promedio global fue 95.42 % de precisión, 43.1 s de tiempo y 4.81 de claridad.

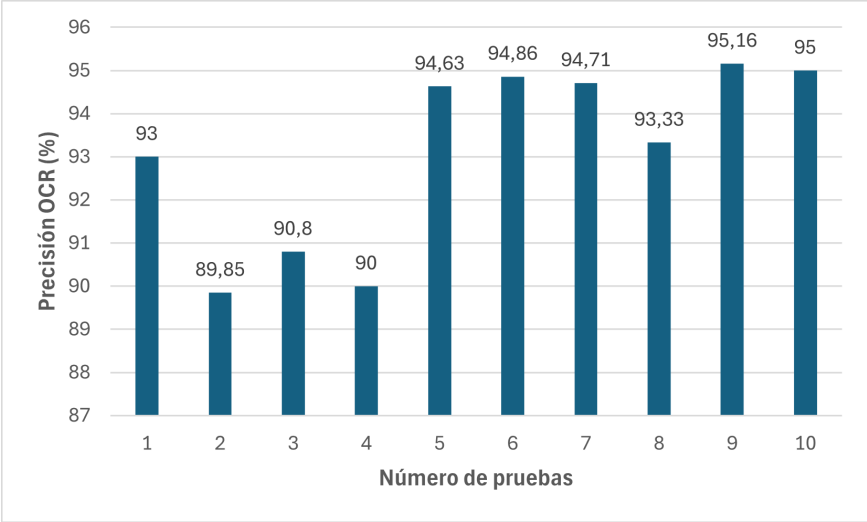
6.3. Iluminación Moderada (222 lux)

6.3.1. Precisión OCR para Fuente 12 pt

El análisis de la Figura 20 revela una disminución notable en la precisión del OCR para la fuente de 12 pt, con un promedio de 93,13 %. Aunque esta configuración enfrenta mayores desafíos en la segmentación de caracteres debido a la iluminación moderada, el sistema logra mantener un desempeño aceptable, indicando su adaptabilidad en entornos donde las condiciones lumínicas no son óptimas y constantes. Aunque la precisión es ligeramente inferior a la de la fuente de 14 pt, sigue siendo suficiente para tareas comunes.

Figura 20

Distribución de precisión OCR para fuente de 12 puntos bajo iluminación moderada.



Análisis de precisión durante las iteraciones de prueba bajo condiciones de iluminación moderada (222 lux), mostrando la adaptabilidad del sistema.

Bajo iluminación moderada (222 lux), la Tabla 3 resume los valores en cada prueba para la fuente 12 pt. El porcentaje de precisión disminuye ligeramente (entre 89.85 % y 95 %) para documentos entre 298 y 369 palabras.

Los tiempos se ubican en 42–47 segundos, y la claridad del audio (4.5–4.7) se mantiene en

rangos aceptables. Aunque se observan más errores de OCR debido a la menor definición de contraste, el desempeño sigue siendo adecuado para tareas de lectura asistida en entornos de oficina o bibliotecas.

Tabla 3

Resultados detallados de OCR (Tamaño de Fuente 12 pt) bajo Iluminación Moderada (222 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	369	343	93.00	47	4.6
2	335	301	89.85	45	4.7
3	336	305	90.80	42	4.5
4	320	288	90.00	46	4.6
5	298	283	94.63	45	4.7
6	350	332	94.86	43	4.65
7	340	323	94.71	44	4.7
8	330	308	93.33	42	4.6
9	310	295	95.16	45	4.5
10	300	285	95.00	44	4.7

Nota: Se obtuvo un promedio de 93.13 % de precisión, 44.3 s y 4.63 de claridad.

6.3.2. Precisión OCR para Fuente 14 pt

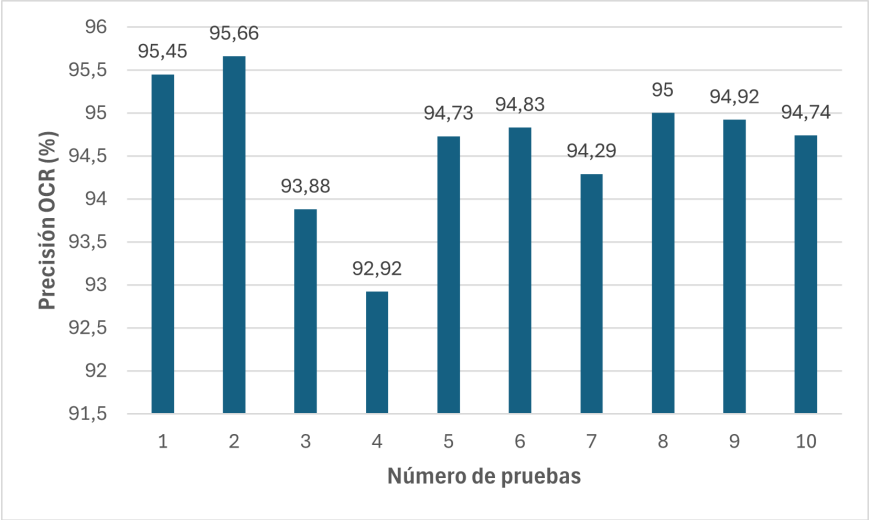
La Figura 21 muestra un desempeño sólido para la fuente de 14 pt en iluminación moderada, con una precisión promedio de 94,64 %. Este resultado destaca la resiliencia del sistema frente a condiciones lumínicas intermedias, donde los caracteres más grandes ofrecen ventajas significativas para el reconocimiento. En comparación con la fuente de 12 pt, se observa una menor susceptibilidad a errores provocados por la disminución en el contraste y la definición de los textos.

Adicionalmente, esta configuración mantiene un nivel de consistencia elevado en las iteraciones de prueba, lo que refuerza su idoneidad para entornos de oficina o iluminación artificial estándar. La

capacidad de procesar textos con mayor claridad y estabilidad posiciona a la fuente de 14 pt como una alternativa confiable para tareas prácticas que requieren adaptabilidad en condiciones lumínicas variadas.

Figura 21

Distribución de precisión OCR para fuente de 14 puntos bajo iluminación moderada.



Evaluación sistemática de la precisión bajo condiciones de iluminación moderada (222 lux).

La Tabla 4 muestra la robustez de la fuente 14 pt cuando la iluminación es moderada. La precisión oscila entre 92.92 % y 95.71 %, lo que supera la fuente de 12 pt bajo estas mismas condiciones.

Aunque el tiempo se mantiene en la banda de 43–46 segundos, no se percibe un gran impacto por la ampliación del tamaño de la fuente. Asimismo, la claridad del audio suele rondar 4.6–4.7, confirmando que la mayor legibilidad del texto reduce los errores OCR que afectarían la pronunciación.

Tabla 4

Resultados detallados de OCR (Tamaño de Fuente 14 pt) bajo Iluminación Moderada (222 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	286	274	95.45	46	4.6
2	277	265	95.66	45	4.7
3	278	261	93.88	44	4.7
4	297	276	92.92	43	4.7
5	283	268	94.73	46	4.6
6	290	275	94.83	45	4.7
7	280	264	94.29	44	4.7
8	300	285	95.00	43	4.7
9	295	280	94.92	45	4.7
10	285	270	94.74	46	4.6

Nota: El promedio general fue 94.64 % de precisión, 44.7 s de tiempo y 4.67 de claridad.

6.4. Iluminación Reducida (10 lux)

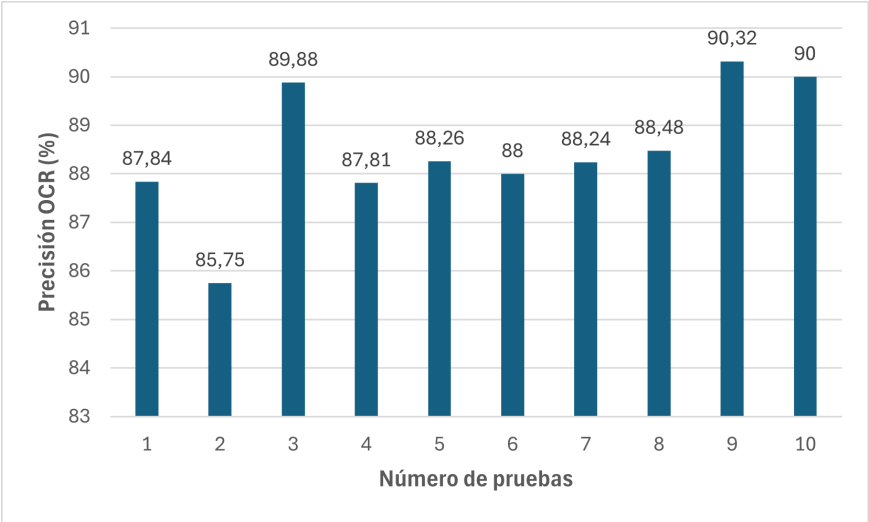
6.4.1. Precisión OCR para Fuente 12 pt

En la Figura 22 se observa una disminución notable en la precisión del OCR para la fuente de 12 pt, con un promedio de 88,46 %. Este descenso refleja las dificultades del sistema para procesar caracteres pequeños en condiciones de baja iluminación, donde el contraste entre texto y fondo es insuficiente para un reconocimiento preciso.

Aunque el desempeño es inferior al de la fuente de 14 pt, sigue siendo útil para documentos en los que el contenido es claramente legible y la iluminación, aunque limitada, no compromete totalmente la calidad visual.

Figura 22

Distribución de precisión OCR para fuente de 12 puntos bajo iluminación reducida.



Análisis detallado bajo iluminación reducida (10 lux), mostrando los límites operativos del sistema.

Bajo condiciones de baja luminosidad (10 lux), la Tabla 5 refleja valores mínimos de 85.75 % (prueba 2) y máximos cercanos al 90.32 % (prueba 9). Este descenso en la exactitud se asocia a la falta de contraste entre texto y fondo, dificultando la segmentación de caracteres pequeños (12 pt).

El tiempo de procesamiento puede incrementarse ligeramente, entre 42 y 48 segundos, debido a que el sistema intenta compensar con varios ajustes de enfoque y un preprocesamiento más cuidadoso. La claridad promedio de 4.48 indica que la síntesis de voz se ve algo afectada por los errores OCR, pero sigue siendo entendible.

Tabla 5

Resultados detallados de OCR (Tamaño de Fuente 12 pt) bajo Iluminación Baja (10 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	369	324	87.84	48	4.4
2	335	288	85.75	46	4.6
3	336	302	89.88	43	4.5
4	320	281	87.81	42	4.4
5	298	262	88.26	47	4.5
6	350	308	88.00	44	4.45
7	340	300	88.24	46	4.55
8	330	292	88.48	43	4.5
9	310	280	90.32	45	4.4
10	300	270	90.00	46	4.5

Nota: El promedio final fue 88.46 % de precisión, 45 s de tiempo y 4.48 en claridad.

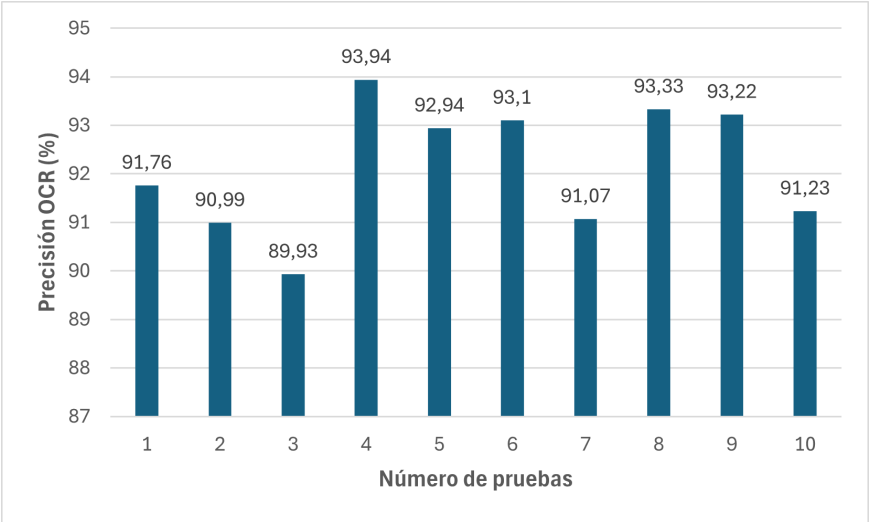
6.4.2. Precisión OCR para Fuente 14 pt

La Figura 23 muestra que, incluso en iluminación reducida, la fuente de 14 pt mantiene un desempeño destacable, con una precisión promedio de 92,15 %. Este resultado evidencia la capacidad del sistema para operar de manera confiable bajo condiciones críticas, donde los caracteres más grandes mitigan los efectos negativos de la baja luminosidad al ofrecer mayor claridad y contraste.

Estas condiciones representan un desafío para el sistema, pero la fuente de 14 pt demuestra ser una opción confiable para ambientes con poca luz, como la lectura de documentos antiguos o materiales en almacenes mal iluminados.

Figura 23

Distribución de precisión OCR para fuente de 14 puntos bajo iluminación reducida.



Evaluación de precisión en iluminación reducida (10 lux), evidenciando la robustez del sistema con caracteres más grandes.

Finalmente, la Tabla 6 comprueba la ventaja de un tamaño de fuente mayor (14 pt) en condiciones críticas (10 lux). Aunque la iluminación es reducida, la precisión promedio (92.15 %) se mantiene relativamente alta, con valores de hasta 93.94 % (prueba 4).

El tiempo (41–48 s) no difiere significativamente de escenarios con más luz, y la claridad del audio (4.4–4.6) evidencia que los errores de reconocimiento impactan menos en la pronunciación, presumiblemente por la mejor detección de caracteres grandes aun con baja luz. Esto sugiere que, para documentos que se leen habitualmente en espacios poco iluminados, la fuente 14 pt sigue siendo viable y produce textos más precisos.

Tabla 6

Resultados detallados de OCR (Tamaño de Fuente 14 pt) bajo Iluminación Baja (10 lux).

Prueba	Núm. Palabras	Palabras Correctas	Precisión OCR (%)	Tiempo (s)	Claridad
1	286	263	91.76	45	4.4
2	277	252	90.99	43	4.5
3	278	250	89.93	48	4.6
4	297	279	93.94	46	4.5
5	283	263	92.94	41	4.4
6	290	270	93.10	44	4.45
7	280	255	91.07	43	4.55
8	300	280	93.33	47	4.5
9	295	275	93.22	45	4.4
10	285	260	91.23	46	4.5

Nota: El promedio general fue 92.15 % de precisión, 44.8 s de tiempo y 4.48 en claridad.

6.5. Análisis Comparativo de Precisión OCR

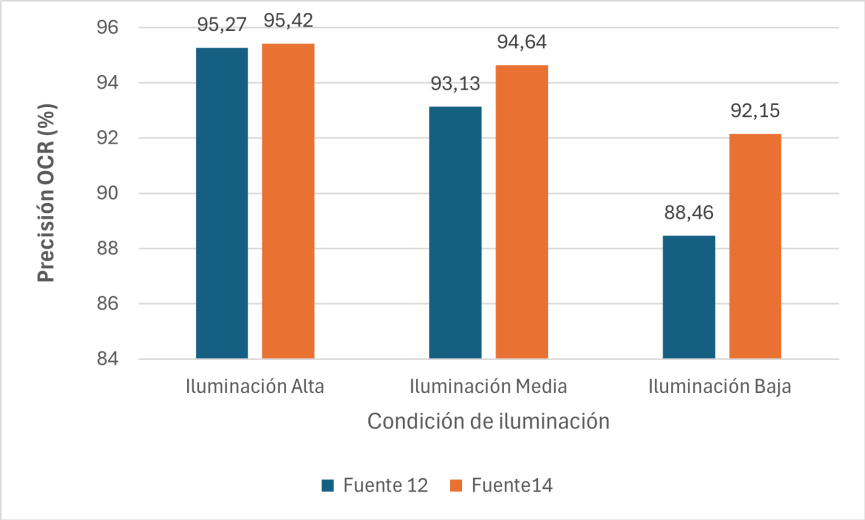
La Figura 24 presenta un análisis comparativo exhaustivo de la precisión OCR promedio entre las fuentes de 12 pt y 14 pt bajo las tres condiciones de iluminación evaluadas. Los resultados evidencian:

- En condiciones de iluminación intensa, ambos tamañaos de fuente demostraron un rendimiento comparable, con una ligera superioridad de la fuente de 14 pt (95,42 % frente a 95,27 %).
- Bajo iluminación moderada, la fuente de 14 pt mantiene una superioridad operativa (94,74 %) en comparación con la fuente de 12 pt (93,13 %), demostrando mayor estabilidad en condiciones subóptimas.

- En entornos de iluminación reducida, la fuente de 14 pt sobresale con ventaja significativa con una precisión promedio de 92,15 % frente al 88,46 % de la fuente de 12 pt, evidenciando su adaptabilidad superior en condiciones lumínicas críticas.

Figura 24

Análisis comparativo de precisión promedio bajo diferentes condiciones de iluminación y dimensiones tipográficas.



Síntesis comparativa de los promedios de precisión obtenidos en las iteraciones de prueba para cada escenario de evaluación, permitiendo una visualización integral del rendimiento del sistema.

6.6. Evaluación de la Calidad Acústica

La Figura 25 presenta los resultados de la calidad acústica del sistema de síntesis de voz (gTTS), evaluada mediante un análisis perceptual. Las valoraciones obtenidas se encuentran dentro del rango de 4,48 a 4,81 en todas las configuraciones experimentales. Estos resultados reflejan la efectividad del sistema en mantener una alta calidad de síntesis de voz bajo diversas condiciones de iluminación y tamaños de fuente, destacando su capacidad para generar un audio claro y comprensible.

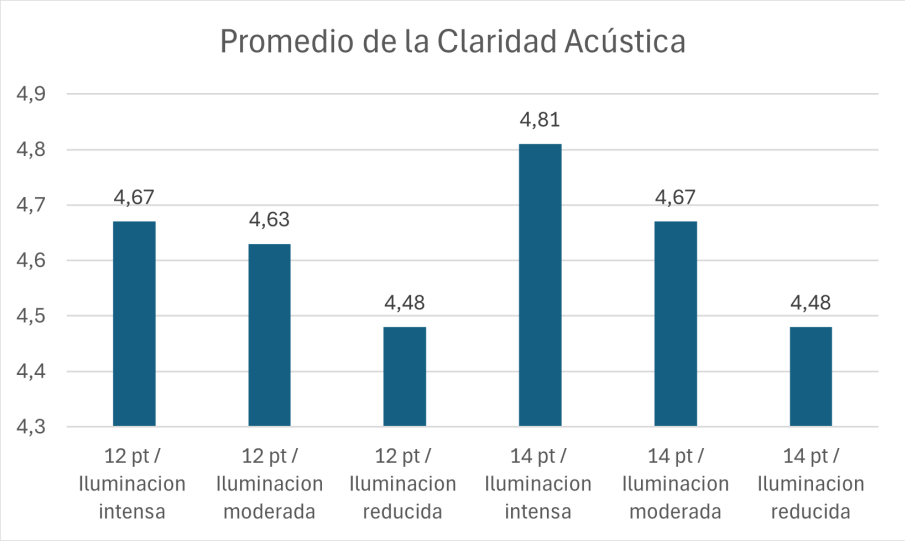
- Para niveles de iluminación alta (1305 lux), la calidad acústica se mantiene en un rango óptimo promedio entre 4,67 y 4,81, evidenciando una correlación positiva entre la luminosidad y la

coherencia en la pronunciación.

- Cuando la iluminocidad desciende a niveles moderados (222 lux) y bajos (10 lux), la calidad acústica experimenta una ligera degradación, con valores entre 4,48 y 4,67, atribuible a condiciones de iluminación menos óptimas que impactan en la fluidez y comprensibilidad del contenido sintetizado.

Figura 25

Promedio de Claridad del Audio en Función de la Iluminocidad.



La claridad del audio varía en función de la iluminocidad y el tamaño de fuente. Se observa que, a medida que la iluminocidad aumenta, la claridad del audio también mejora.

6.7. Análisis del Tiempo de Procesamiento

La evaluación temporal del procesamiento y rendimiento del sistema en todas sus etapas reveló una consistencia notable, con tiempos que oscilan entre 40 y 48 segundos con una promedio de 44 segundos para documentos de 200 a 400 palabras, independientemente de las variables tipográficas o condiciones de iluminación. Este rango estable demuestra que la implementación es eficiente y predecible, garantizando que no existan latencias significativas durante su operación.

En escenarios prácticos, como la digitalización de documentos en oficinas o bibliotecas, estos tiempos resultan aceptables para tareas de lectura asistida. Para para un documento promedio de 300 palabras, el sistema es capaz de procesar el texto en menos de un minuto, asegurando una experiencia de usuario fluida.

6.8. Discusión de Resultados

El análisis de los datos demuestran que la cantidad de luz es un factor clave que influye en el funcionamiento del sistema OCR y, en consecuencia, en la claridad de la voz generada.

Bajo condiciones de iluminación alta (1305 lux), las fuentes de 12 y 14 puntos logran excelentes resultados, con un promedio de precisión del 95,27 % y 95,42 %, respectivamente, lo que demuestra una eficiencia estable cuando las condiciones de luz son ideales. Además, la calidad de la voz generada obtiene calificaciones entre 4,67 y 4,81 sobre 5, lo que refleja una buena entonación y claridad gracias a la alta precisión del reconocimiento de texto.

Cuando la iluminación disminuye a niveles moderados (222 lux), se observa una ligera caída en el desempeño del sistema. La fuente de 12 puntos logra un promedio de precisión del 93,13 %, mientras que la de 14 puntos alcanza un mejor rendimiento con un 94,64 %. Sin embargo, hay pequeños errores al identificar signos como tildes y palabras técnicas, lo que afecta ligeramente la calidad de la voz, que varía entre 4,63 y 4,67.

Con iluminación baja (10 lux), la precisión del sistema disminuye significativamente. La fuente de 12 puntos alcanza un promedio de 88,4 % de precisión, mientras que la de 14 puntos demuestra ser más resistente con un 92,15 %. La calidad de la voz cae a 4,48 sobre 5, reflejando las dificultades en el reconocimiento de texto en estas condiciones. Esto sugiere que hay un límite mínimo de precisión necesario para que la voz generada sea clara y entendible.

En términos generales, el tiempo que el sistema tarda en procesar los documentos se mantiene por debajo de los 44 segundos, lo cual es adecuado para textos de longitud moderada (entre 200 y 400 palabras).

7. CONCLUSIONES, RECOMENDACIONES Y TRABAJOS A FUTURO

7.1. Conclusiones

1. Tras la validación de los requerimientos de hardware y software, se concluye que la Raspberry Pi 4 ofrece un equilibrio adecuado entre costo y rendimiento para ejecutar OpenCV, Tesseract y gTTS. Su arquitectura compacta y el soporte de múltiples librerías permiten desarrollar sistemas de lectura asistida de bajo costo, manteniendo niveles aceptables de eficiencia y versatilidad.

2. La detección de texto se implementó con el modelo EAST, basado en técnicas de Deep Learning, y la extracción OCR con Tesseract, complementada con SpaCy para la corrección semántica. Esta integración modular demostró solidez a partir de los 222 lux, facilitando una conversión de texto a voz precisa y fluida en condiciones de iluminación moderadas a altas.

3. En ambientes bien iluminados ($\geq$ 222 lux), el sistema alcanzó un 95 % de precisión, procesando de forma rápida (44 s) documentos de 200–400 palabras. Sin embargo, en escenarios de baja luminosidad (10 lux) y con fuentes pequeñas (12 pt), la precisión descendió a aproximadamente el 88 %, evidenciando la necesidad de optimizar el preprocesamiento y garantizar una iluminación suficiente para evitar pérdidas significativas en el reconocimiento.

4. El modelo EAST, basado en redes neuronales convolucionales, demostró ser efectivo para localizar texto incluso en diversos tamaños y orientaciones. No obstante, su sensibilidad a la falta de contraste implica que se necesite iluminación adecuada para obtener resultados óptimos. Ante escenarios con luz muy reducida, el detector puede omitir regiones importantes o generar bounding boxes imprecisos.

5. El reconocimiento de caracteres mediante Tesseract, respaldado por redes neuronales LSTM, mostró resultados fiables en idioma español. Aun así, la presencia de imágenes con baja

nitidez o luminancia provoca un incremento en la tasa de errores, subrayando la importancia de una buena calidad de imagen (determinada por la cámara y la iluminación) para mantener un porcentaje de acierto elevado.

7.2. Recomendaciones

En primer lugar, es fundamental mejorar la captura de imágenes mediante la incorporación de fuentes de iluminación adicionales, como focos LED o lámparas ajustables, que puedan adaptarse a diferentes entornos y reducir las sombras o reflejos que afectan la precisión del OCR. Además, optimizar el preprocesamiento de imágenes mediante técnicas avanzadas de aumento de contraste y reducción de ruido, facilitará una mejor detección y reconocimiento de texto incluso en condiciones de iluminación adversas.

Otra recomendación es ampliar los diccionarios y modelos utilizados por Tesseract, incluyendo vocabularios especializados que abarcan nombres propios, términos científicos y técnicos. Esto permitiría una mayor exactitud en el reconocimiento de palabras específicas, mejorando así la calidad del texto sintetizado por gTTS.

7.3. Trabajos a Futuro

Una de las principales líneas de mejora consiste en la optimización del pipeline de visión, explorando modelos de detección y reconocimiento de texto más avanzados, como DBNet o CRNN, que ofrecen una mayor tolerancia a fondos complejos y variaciones en la iluminación. Además, la incorporación de aceleradores de hardware, como Coral USB Accelerator, podría reducir considerablemente los tiempos de procesamiento, mejorando la eficiencia del sistema en dispositivos de bajo costo como la Raspberry Pi.

Extender el sistema a la lectura de manuscritos, lo que representaría un desafío adicional para los motores OCR convencionales. Esto ampliaría la utilidad del sistema en contextos educativos y laborales, donde la lectura de texto manuscrito es frecuente. Además, la integración de módulos de traducción automática, utilizando servicios como Google Translate, permitiría manejar textos en

múltiples idiomas, aumentando la versatilidad del sistema en entornos multilingües y facilitando su uso en contextos académicos o turísticos.

Asimismo, se contempla la portabilidad del sistema a plataformas móviles, desarrollando aplicaciones para Android e iOS que permitan a los usuarios aprovechar las cámaras de sus teléfonos para capturar y convertir texto a voz en cualquier momento y lugar. Esta portabilidad incrementaría la accesibilidad y adopción del sistema, haciéndolo más conveniente para los usuarios finales.

Referencias

AI, E. (2024). spaCy: Industrial-strength Natural Language Processing in Python. https://spacy.io/

Ani, R., Maria, E., Joyce, J. J., Sakkaravarthy, V., & Raja, M. A. (2017). Smart Specs: Voice assisted text reading system for visually impaired persons using TTS method. *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, 1-6. https://doi.org/10.1109/IGEHT.2017.8094103

Antonacopoulos, A., & Bridson, D. (2004). A Robust Braille Recognition System. En S. Marinai & A. R. Dengel (Eds.), *Document Analysis Systems VI* (pp. 533-545). Springer Berlin Heidelberg.

Beinat, N. C. (2022). *Redes Neuronales Convolucionales y Aplicaciones*. https://eprints.ucm.es/id/eprint/70850/

Chang, A. C.-S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, *52*, 91-102. https://doi.org/https://doi.org/10.1016/j.system.2015.05.003

Chaudhuri, A., Mandaviya, K., Badelia, P., & Ghosh, S. (2017, diciembre). Optical Character Recognition Systems. En *Optical Character Recognition Systems for Different Languages with Soft Computing* (pp. 9-41, Vol. 352). Springer. https://doi.org/10.1007/978-3-319-50252-6_2

Cheriet, M., Kharma, N., Liu, C.-L., & Suen, C. Y. (2007). Tools for Image Preprocessing. En C. Y. Suen (Ed.), *Character Recognition Systems: A Guide for Students and Practitioners* (pp. 5-53). John Wiley & Sons. https://doi.org/10.1002/9780470176535.ch2

Consejo Nacional para la Igualdad de Discapacidades. (2024). Estadísticas de Discapacidad. https://www.consejodiscapacidades.gob.ec/estadisticas-de-discapacidad/

Culjak, I., Abram, D., Pribanic, T., Dzapo, H., & Cifrek, M. (2012). A brief introduction to OpenCV. *2012 Proceedings of the 35th International Convention MIPRO*, 1725-1730.

Curiel Centenero, C., & et al. (2018). *Impacto de la Calidad de Vida en Niños con Discapacidad Visual y Manejo en el Aula* [Trabajo de Fin de Máster]. Universidad de Valladolid. https://uvadoc.uva.es/handle/10324/31773

Elyan, E., Vuttipittayamongkol, P., Johnston, P., Martin, K., McPherson, K., Moreno-García, C. F., Jayne, C., & Sarker, M. M. K. (2022). Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. *Artificial Intelligence Surgery*, *2*(1). https://doi.org/10.20517/ais.2021.15

Fichten, C. S., Asuncion, J. V., Barile, M., Ferraro, V., & Wolforth, J. (2009). Accessibility of e-Learning and Computer and Information Technologies for Students with Visual Impairments in Postsecondary Education. *Journal of Visual Impairment & Blindness*, *103*(9), 543-557. https://doi.org/10.1177/0145482X0910300905

Foundation, P. S. (2021). *Python: A Powerful and Versatile Programming Language*. https://www.python.org/

Genius. (2020). Genius SP-U115 Speakers [Consultado el 2 de diciembre de 2024]. https://www.geniusnet.com/product/sp-u115/

Hamad, K., & Kaya, M. (2016). A Detailed Analysis of Optical Character Recognition Technology. *International Journal of Applied Mathematics, Electronics and Computers*, *4*, 244-244. https://doi.org/10.18100/ijamec.270374

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huynh, A., Marchant, Z., Singh, I., & Aveta, F. (2023). Assistive Reading Device for Blind Individuals. *2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0606-0612. https://doi.org/10.1109/UEMCON59035.2023.10316082

JUGRAN, S., KUMAR, A., TYAGI, B. S., & ANAND, V. (2021). Extractive Automatic Text Summarization using SpaCy in Python & NLP. *2021 International Conference on Advance*

Computing and Innovative Technologies in Engineering (ICACITE), 582-585. https://doi.org/10.1109/ICACITE51222.2021.9404712

Karmel, A., Sharma, A., pandya, M., & Garg, D. (2019). IoT based Assistive Device for Deaf, Dumb and Blind People [2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019]. *Procedia Computer Science*, *165*, 259-269. https://doi.org/https://doi.org/10.1016/j.procs.2020.01.080

Kashyap, R., & Kumar, A. V. S. (Eds.). (2019). *Challenges and Applications for Implementing Machine Learning in Computer Vision*. IGI Global. https://doi.org/10.4018/978-1-7998-0182-5

Khan, A. I., & Al-Habsi, S. (2020). Machine Learning in Computer Vision [International Conference on Computational Intelligence and Data Science]. *Procedia Computer Science*, *167*, 1444-1451. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.355

Khan, M. A., Paul, P., Rashid, M., Hossain, M., & Ahad, M. A. R. (2020). An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind. *IEEE Transactions on Human-Machine Systems*, *50*(6), 507-517. https://doi.org/10.1109/THMS.2020.3027534

Leporini, B., & Paternò, F. (2004). Increasing usability when interacting through screen readers. *Universal Access in the Information Society*, *3*(1), 57-70. https://doi.org/10.1007/s10209-003-0076-4

Lin, H.-Y. (2022). Large-Scale Artificial Intelligence Models. *Computer*, *55*(5), 76-80. https://doi.org/10.1109/MC.2022.3151419

Logitech. (2019). Logitech C920s HD Pro Webcam [Consultado el 2 de diciembre de 2024]. https://www.logitech.com/en-us/products/webcams/c920s-pro-hd-webcam.960-001257.html

Microsoft Corporation. (s.f.). Seeing AI. https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/

Ministerio de Salud Pública de Ecuador. (2022). Ecuador avanza hacia un proceso inclusivo y de reducción de las desigualdades para personas con discapacidad. https://www.salud.gob.ec/

ecuador-avanza-hacia-un-proceso-inclusivo-y-de-reduccion-de-las-desigualdades-para-personas-con-discapacidad/

Najam, R., & Faizullah, S. (2023). Analysis of Recent Deep Learning Techniques for Arabic Handwritten-Text OCR and Post-OCR Correction. *Applied Sciences*, *13*(13). https://doi.org/10.3390/app13137568

OrCam Staff. (2022, abril). La tecnología al servicio de las personas con discapacidad visual. https://www.orcam.com/es-es/blog/la-tecnologia-al-servicio-de-las-personas-con-discapacidad-visual

Organización Mundial de la Salud. (s.f.). Clasificación Internacional de Enfermedades (CIE). https://www.who.int/classifications/classification-of-diseases

Organización Mundial de la Salud. (2019). La OMS presenta el primer Informe mundial sobre la visión. https://www.who.int/es/news/item/08-10-2019-who-launches-first-world-report-on-vision

Organización Panamericana de la Salud. (s.f.). Salud visual. https://www.paho.org/es/temas/salud-visual

Pallero, R. (2008). Ajuste psicosocial a la discapacidad visual en personas mayores. *Integración: Revista sobre ceguera y deficiencia visual, ISSN 0214-1892, Nº. 55, 2008 (Ejemplar dedicado a: Envejecimiento y discapacidad visual), pags. 34-42*, *55*.

Palmér, M. (2024). pynput: Monitor and Control Input Devices. https://pypi.org/project/pynput/

Praveen, P., & Madihabanu, S. (2023). A Real Time Multiple Object Tracking in Videos using CNN Algorithm. *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 1-6. https://doi.org/10.1109/ICSSAS57918.2023.10331876

Pujante, B. G. (2023). *Redes Convolucionales. Aplicación a la clasificación de imágenes médicas*. https://dspace.umh.es/handle/11000/30233

Rahmati, M., Fateh, M., Rezvani, M., Tajary, A., & Abolghasemi, V. (2020). Printed Persian OCR system using deep learning. *IET Image Processing*, *14*(15), 3920-3931. https://doi.org/https://doi.org/10.1049/iet-ipr.2019.0728

Ranjan, A., Behera, V. N. J., & Reza, M. (2021). OCR Using Computer Vision and Machine Learning. En S. Das, S. Das, N. Dey & A. E. Hassanien (Eds.), *Machine Learning Algorithms for Industrial Applications* (pp. 101-117, Vol. 907). Springer. https://doi.org/10.1007/978-3-030-50641-4_6

Rasinski, T., & Young, C. (2014). Assisted reading–A bridge from fluency to comprehension. *New England Reading Association Journal*, *50*(1), 1-4.

Raspberry Pi Foundation. (2019). *Raspberry Pi 4 Model B* [Consultado el 2 de diciembre de 2024]. https://www.raspberrypi.com/products/raspberry-pi-4-model-b/

Sangpal, R., Gawand, T., Vaykar, S., & Madhavi, N. (2019). JARVIS: An interpretation of AIML with integration of gTTS and Python. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, *1*, 486-489. https://doi.org/10.1109/ICICICT46008.2019.8993344

Sarwar, S., Turab, M., Channa, D., Chandio, A., Sohu, M. U., & Kumar, V. (2022). Advanced Audio Aid for Blind People. *2022 International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC)*, 1-6. https://doi.org/10.1109/ICETECC56662.2022.10069052

Smith, R. (2007). An Overview of the Tesseract OCR Engine. *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, *2*, 629-633. https://doi.org/10.1109/ICDAR.2007.4376991

Smith, R., et al. (2024). Tesseract OCR. https://github.com/tesseract-ocr/tesseract

Sociedad Panamericana de Retinopatía del Prematuro. (2020). Estadísticas. https://sprop.org/estadisticas/

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding Long Short-Term Memory Recurrent Neural Networks – a tutorial-like introduction. *arXiv preprint arXiv:1909.09586*. https://arxiv.org/abs/1909.09586

Suma, V. (2019). Computer vision for human-machine interaction-review. *Journal of trends in Computer Science and Smart technology (TCSST)*, *1*(02), 131-139.

Surana, S., Pathak, K., Gagnani, M., Shrivastava, V., R, M. T., & Madhuri G, S. (2022). Text Extraction and Detection from Images using Machine Learning Techniques: A Research Review. *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, 1201-1207. https://doi.org/10.1109/ICEARS53579.2022.9752274

Team, O. (2024). OpenCV: Open Source Computer Vision Library. https://opencv.org/

Techly. (2023). Mini Numeric Keypad with USB Cable and 18 Keys. https://www.techly.com/mini-numeric-keypad-with-usb-cable-and-18-keys.html

The international agency for the prevention of blindness. (2021). Catarata como causa de ceguera y discapacidad visual en América Latina: avances logrados y retos después de 2020. https://www.iapb.org/news/catarata-como-causa-de-ceguera-y-discapacidad-visual-en-america-latina-avances-logrados-y-retos-despues-2020/

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. https://arxiv.org/abs/1609.03499

VideoLAN. (2024). *VLC Media Player*. https://www.videolan.org/vlc/

Yellapu, V., S, R., B, H., Sharma, S., & Dev, M. (2022). Development of Communication System For Deaf And Blind Persons Using Text to Braille Conversion. *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, 69-73. https://doi.org/10.1109/IIHC55949.2022.10060434

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: An Efficient and Accurate Scene Text Detector. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2642-2651. https://doi.org/10.1109/CVPR.2017.283

ANEXOS

7.4. Dispositivo Final Montado

En la Figura 26, se observa el montaje final del dispositivo de lectura asistida. El diseño del soporte facilita una posición fija y estable para la cámara, permitiendo una captura de imagen consistente y precisa. Además, el marco para la hoja de texto está diseñado como guía para la colocación del documento a procesar.

Figura 26

Vista general del dispositivo final montado en un soporte, incluyendo el adaptador para la cámara y un marco diseñado específicamente para sostener la hoja de texto.

La imagen muestra el dispositivo completo ensamblado en su estructura de soporte.