

DISEÑO E IMPLEMENTACIÓN DE UN DISPOSITIVO ELECTRÓNICO PARA ASISTENCIA DE LECTURA DE TEXTOS IMPRESOS MEDIANTE VISIÓN COMPUTACIONAL PARA PERSONAS CON DISCAPACIDAD VISUAL

CRISTHIAN KEVIN RAFAEL ROSALES*

<https://orcid.org/0009-0004-8642-0627>

Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería, Perú

JOSE ERASMO SALAZAR MINCHAN

<https://orcid.org/0009-0004-4791-1466>

Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería, Perú

JORGE ENRIQUE ORTIZ PORRAS

<https://orcid.org/0000-0002-9605-3670>

Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería, Perú

Recibido: 6 de febrero del 2025 / Aceptado: 10 de marzo del 2025

Publicado: 10 de junio del 2025

doi: <https://doi.org/10.26439/ing.ind2025.n48.7607>

RESUMEN. Este artículo presenta el diseño e implementación de un dispositivo electrónico para ayudar a personas con discapacidades visuales en la lectura de textos impresos. Este sistema mecatrónico utiliza una cámara Raspberry Pi V3 montada en unas gafas para capturar imágenes de textos impresos. Las imágenes se procesan mediante el reconocimiento óptico de caracteres, y el texto extraído se convierte a audio mediante un sintetizador de texto a voz, lo que permite al usuario acceder al contenido visual de manera auditiva. El dispositivo se basa en un microcontrolador Raspberry Pi 3 modelo B+ para la gestión de datos y el control de los componentes elec-

Este estudio no fue financiado por ninguna entidad.

* Autor corresponsal

Correos electrónicos en orden de aparición: crafaelr@uni.pe; jose.salazar.m@uni.pe; jortizp@uni.edu.pe

Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Attribution 4.0 International (CC BY 4.0).

trónicos. Está diseñado para ser portátil, ergonómico y de bajo consumo energético, con una batería de larga duración.

PALABRAS CLAVE: Raspberry Pi / reconocimiento óptico de caracteres / texto a voz / discapacidad visual

DESIGN AND IMPLEMENTATION OF AN ELECTRONIC DEVICE FOR READING ASSISTANCE OF PRINTED TEXTS USING COMPUTER VISION FOR VISUALLY IMPAIRED PEOPLE

ABSTRACT. This article presents the design and implementation of an electronic device to assist visually impaired people in reading printed texts. This mechatronic system uses a Raspberry Pi V3 camera mounted on glasses to capture images of printed texts. The images are processed through Optical Character Recognition, and the extracted text is converted to audio using a Text-to-Speech synthesizer, allowing the user to access the visual content audibly. The device is based on a Raspberry Pi 3 Model B+ microcontroller for data management and control of the electronic components. It is designed to be portable, ergonomic, and energy-efficient, with a long-lasting battery.

KEYWORDS: Raspberry Pi / optical character recognition / text to speech / visually impaired

INTRODUCCIÓN

El acceso a la información es esencial para el desarrollo integral de las personas con discapacidad visual, ya que les permite ejercer su autonomía, participar activamente en la sociedad y acceder a oportunidades de aprendizaje y empleo (Yataco Marín, 2022). Sin embargo, las personas con discapacidad visual enfrentan diversas dificultades para interpretar el texto impreso debido a problemas de alineación, enfoque, precisión, movilidad y eficiencia (Manage et al., 2020). Según Wong Cam et al. (2019), esta condición se define como la dificultad de una persona para realizar actividades cotidianas debido a la pérdida de visión causada por una afección ocular con una duración superior a seis meses.

La Organización Mundial de la Salud (2023) estima que más de 2 200 millones de personas en el mundo tienen alguna forma de discapacidad visual, cuyas principales causas son los errores de refracción no corregidos y las cataratas no tratadas. En el Perú, 1 550 196 personas presentan discapacidad visual, según el perfil sociodemográfico de la población con discapacidad (Instituto Nacional de Estadística e Informática, 2019). Además, hasta enero del 2025, el Consejo Nacional para la Integración de la Persona con Discapacidad (2025), mediante el Registro Nacional de la Persona con Discapacidad (RNPCD), ha inscrito a 542 905 personas con discapacidad, de las cuales 466 835 presentan discapacidad visual.

La tecnología ha facilitado la vida cotidiana, pero las personas con discapacidades físicas aún enfrentan numerosas dificultades (Karmel et al., 2019). En este contexto, la revisión de Kuriakose et al. (2022) destacó que muchos dispositivos asistivos presentan limitaciones en cuanto a usabilidad, portabilidad y costo, lo que puede contribuir a la exclusión social y restringir la autonomía de las personas con discapacidad visual.

Aunque estas barreras persisten, diversas investigaciones han desarrollado soluciones enfocadas en mejorar el acceso a la información para personas con discapacidad visual, principalmente a través del reconocimiento óptico de caracteres (*optical character recognition*, OCR) y la conversión de texto a voz (*text to speech*, TTS). Entre estas soluciones destacan dispositivos como las gafas inteligentes (López Jiménez, 2019), consideradas como un avance en ayudas electrónicas, que integran un procesamiento de imágenes y realidad aumentada para asistir en la lectura y control de diversos factores visuales, lo que busca la autonomía del usuario. A continuación, se presentan algunos antecedentes relevantes en este campo.

Basantes Varela y Chalaco Chamba (2019) presentaron un dispositivo capaz de capturar imágenes de documentos impresos, procesarlas mediante visión artificial y convertir el texto en audio. Las pruebas realizadas demostraron que la distancia mínima de reconocimiento fue de 20 cm y que, para todos los usuarios, la claridad

del audio fue suficiente para comprender el texto. Sin embargo, el prototipo presenta algunas limitaciones, como la restricción en la distancia de detección, la necesidad de marcar los bordes de las hojas con contornos de color verde para su identificación y dificultades con ciertos tipos de letra y tamaños de texto impreso.

Luna Soria (2022) desarrolló un dispositivo estático que permite la inserción de documentos de texto para su digitalización y conversión a audio, cuyo sistema emplea OpenCV para el reconocimiento óptico de caracteres y la API de Google Cloud TTS para la síntesis de voz. En las pruebas finales, el dispositivo alcanzó una eficiencia superior al 95 % en la conversión de texto a audio. Sin embargo, su tamaño considerable limitó su portabilidad, lo que dificultó su uso en entornos cotidianos. Además, al depender de la API de Google Cloud TTS, requiere una conexión a internet para su funcionamiento, lo que podría representar una desventaja en contextos de conectividad limitada.

Arias Acosta (2022) propuso un dispositivo portátil para una lectura audible de textos mediante la implementación de la API de Google Cloud Vision para el OCR y la API de Google Text-to-Speech (gTTS) para la conversión de texto a voz. Las pruebas demostraron una eficiencia del 95,94 % en la detección de textos. Sin embargo, el tamaño de la carcasa que albergaba los componentes electrónicos seguía siendo considerable, lo que limitó su portabilidad para el uso diario o en entornos domésticos. Además, su funcionalidad dependía de una conexión a internet, lo que representó una desventaja en áreas con acceso limitado a la red.

AlSaid et al. (2019) desarrollaron unas gafas inteligentes para asistir a estudiantes con discapacidad visual en la lectura de textos impresos. El sistema utilizó Tesseract OCR y el detector EAST para extraer texto de imágenes capturadas por una cámara integrada, mientras que la conversión de texto a voz se realizó mediante gTTS. Además, incorporaron un sensor RFID para ayudar a los usuarios a ubicarse dentro de un entorno universitario.

Navarro Pérez (2021) desarrolló un dispositivo portátil y de bajo costo diseñado para asistir a personas con discapacidad visual. Este dispositivo integra una Raspberry Pi Zero y una cámara para capturar imágenes de texto impreso. Las imágenes capturadas se procesan mediante OpenCV y se reconocen con Tesseract OCR, para luego convertir el texto en voz a través de un sistema TTS (Pico TTS). Los resultados demostraron que esta herramienta compacta y accesible es capaz de reproducir en audio en diversos tipos de texto, con un nivel de precisión del 88 % al 90 %. Sin embargo, esta precisión, aunque funcional, podría considerarse relativamente baja para aplicaciones de asistencia visual, donde la exactitud es fundamental.

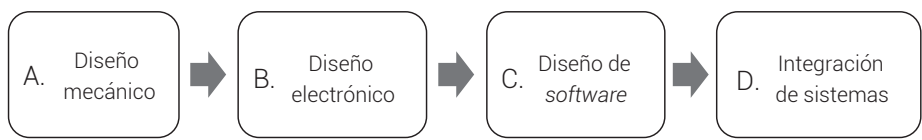
Considerando las limitaciones de estudios previos, este artículo propone un dispositivo electrónico acoplado a gafas que facilite la lectura de textos impresos

mediante visión computacional. El sistema integra una cámara para capturar imágenes, un módulo OCR para convertirlas en texto y un sintetizador TTS para su conversión en audio, lo que permite al usuario acceder al contenido de forma auditiva. A diferencia de otros dispositivos que requieren conexión a internet, este funciona de manera autónoma y garantiza una mayor accesibilidad en distintos entornos. Además, se diseñó una carcasa compacta para asegurar la portabilidad y discreción, lo que facilita su integración en la vida diaria.

METODOLOGÍA

Para abordar de manera sistemática y eficiente el desarrollo del dispositivo con OCR y TTS, se ha dividido el proyecto en las siguientes etapas, como se muestra en la Figura 1:

Figura 1
Diagrama de bloques



Diseño mecánico

En esta etapa, se llevó a cabo el modelado 3D de los componentes mecánicos del dispositivo mediante SolidWorks, lo que permitió visualizar, diseñar y verificar cada pieza antes de su fabricación. Para la elaboración del prototipo, se seleccionó PLA por su resistencia mecánica, accesibilidad y estabilidad estructural en aplicaciones funcionales. Además, su alta resistencia a la fractura lo hace un material idóneo para la fabricación aditiva, pues garantiza durabilidad y fiabilidad en el diseño del dispositivo (Cuesta et al., 2019).

A continuación, en la Tabla 1 se presentan las principales propiedades mecánicas del material PLA, las cuales fueron utilizadas para la simulación en Ansys.

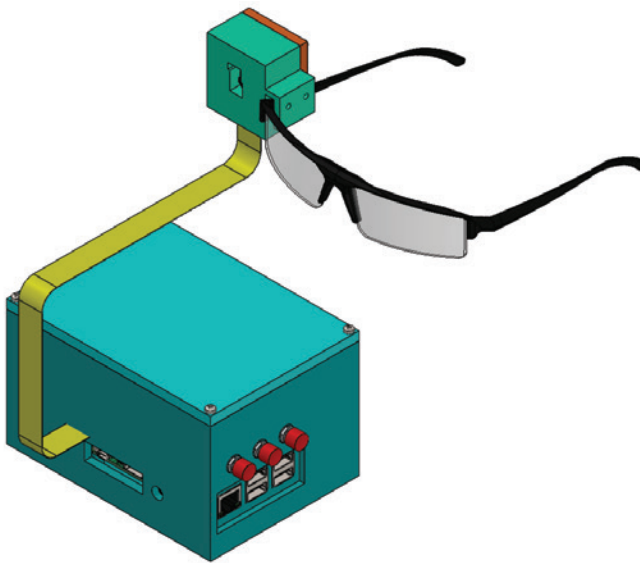
Tabla 1
Propiedades mecánicas del material PLA

Propiedades generales	Valores
Módulo de Young	3,5 GPa
Coefficiente de Poisson	0,35
Densidad	1250 kg/m³

En la Figura 2, se presenta el prototipo del dispositivo. En la parte inferior, se encuentra una carcasa que alberga todos los componentes, incluyendo la Raspberry Pi, la batería y las conexiones necesarias. En la parte superior, se ubica la carcasa de la cámara, que está diseñada para ajustarse a las gafas de manera segura y funcional. Este diseño asegura que todos los componentes estén protegidos y que el dispositivo sea fácil de usar para personas con discapacidades visuales.

Figura 2

Prototipo del dispositivo



Diseño electrónico

En esta etapa del proyecto, se llevó a cabo una evaluación exhaustiva de las alternativas disponibles para los diferentes componentes electrónicos del dispositivo. El objetivo fue seleccionar los mejores componentes en función de criterios como el rendimiento, la compatibilidad, la eficiencia energética y el costo. Los principales aspectos evaluados y seleccionados fueron los siguientes:

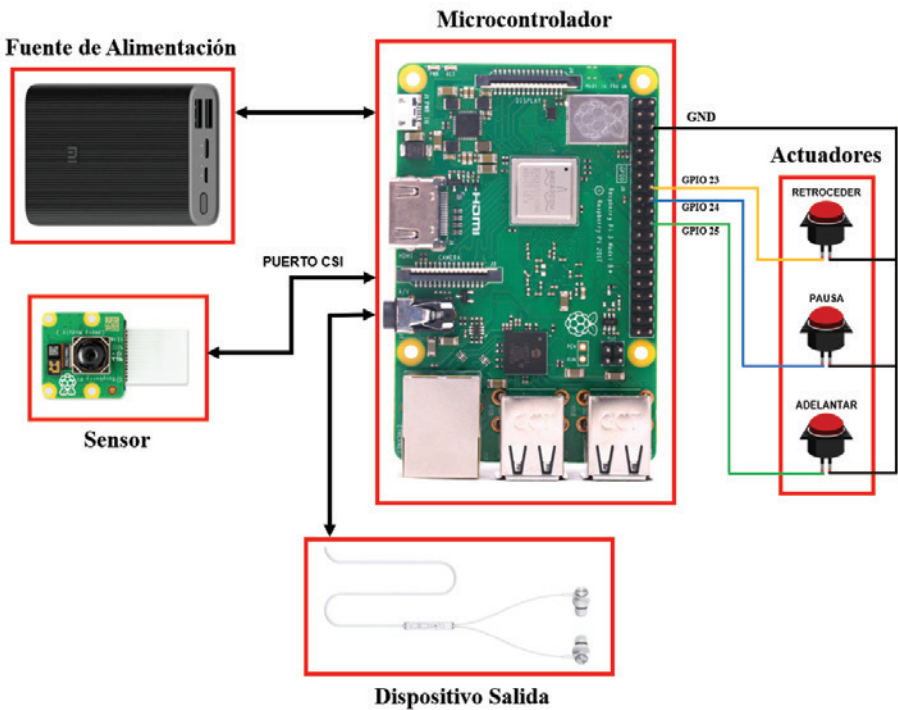
- *Selección del microcontrolador.* Tras evaluar tres alternativas, se eligió la Raspberry Pi 3 Modelo B+ debido a su equilibrio entre rendimiento y costo. Según Raspberry Pi Foundation (2023), esta placa está equipada con un procesador de cuatro núcleos a 1,4 GHz y 1 GB de memoria RAM, lo cual ofrece la potencia y la memoria necesarias para ejecutar tareas complejas, como el reconocimiento de imágenes y la conversión de texto a voz.

- *Selección de la cámara.* Tras evaluar dos opciones, se optó por la cámara Raspberry Pi V3, un módulo especializado que ofrece una resolución de 12 megapíxeles (4608×2592 píxeles) para capturas de alta calidad. Según Raspberry Pi Foundation (2024), esta cámara es compatible con la Raspberry Pi 3, incluye enfoque automático y un sensor optimizado para condiciones de baja luz, lo que asegura imágenes nítidas y detalladas. Estas características son esenciales para la fase de reconocimiento de texto, pues permiten una interpretación precisa de los caracteres y mejoran la precisión del sistema.
- *Selección de la salida de audio.* Se seleccionaron auriculares alámbricos para la salida de audio debido a su fácil integración con la Raspberry Pi y su buena calidad de sonido. Esta elección simplifica el diseño al evitar módulos inalámbricos y configuraciones complejas. Además, son adecuados para entornos silenciosos como bibliotecas, ya que permiten una experiencia auditiva discreta sin afectar al entorno.
- *Selección del sistema de alimentación.* Para la fuente de energía, se eligió el Xiaomi Mi Power Bank 3 Ultra Compact con una capacidad de 10 000 mAh, que proporciona la autonomía necesaria para un uso prolongado sin recargas frecuentes. Su diseño compacto facilita la portabilidad, ideal para un dispositivo que requiere movilidad. En la selección se consideraron tanto su alta capacidad en relación con su tamaño reducido, adecuado para un dispositivo con espacio limitado, como sus sistemas de protección de batería LiPo, esenciales para un funcionamiento seguro y confiable debido a la sensibilidad de esta tecnología (Xiaomi, s. f.).
- *Diagrama de conexión.* La Figura 3 muestra el diagrama de conexión del sistema, en el que se integran los principales componentes: la Raspberry Pi 3 Modelo B+, la cámara Raspberry Pi Cam v3, los auriculares y la batería portátil. La Raspberry Pi actúa como el centro de control, pues gestiona los datos.

La cámara se conecta al puerto CSI para capturar imágenes que serán procesadas mediante OCR, mientras que los auriculares se conectan al puerto de audio para la salida de voz generada por el sistema TTS. Por su parte, la batería externa suministra energía a todo el sistema, lo que garantiza su portabilidad.

Este diagrama es fundamental para comprender y garantizar la correcta integración de los componentes, porque facilita su implementación y verificación.

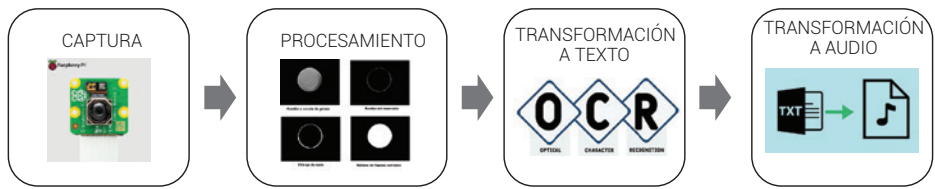
Figura 3
Diagrama de conexión



Diseño del software

Para el desarrollo del sistema de asistencia de lectura mediante reconocimiento óptico de caracteres y conversión de texto a voz, se ha diseñado el diagrama de bloques mostrado en la Figura 4. Este diagrama representa la estructura general del *software* y define los componentes principales del sistema, así como su flujo de información.

Figura 4
Diagrama de bloques del sistema



Los bloques principales incluyen los siguientes módulos:

- *Captura de imagen.* En este módulo, la cámara del dispositivo captura la imagen del texto que el usuario desea leer y la almacena temporalmente en el sistema para su posterior procesamiento.
- *Procesamiento de imagen.* En esta etapa, el *software* utiliza técnicas de preprocesamiento mediante OpenCV (Howse et al., 2020) para optimizar la calidad de la imagen, de modo que ajusta el contraste y reduce el ruido. Estas mejoras permiten destacar los caracteres y preparar la imagen para un reconocimiento de texto más preciso en el sistema OCR.
- *Reconocimiento óptico de caracteres.* En esta fase, el texto se extrae de la imagen mediante un modelo de OCR basado en la biblioteca Tesseract OCR (Smith, 2007), seleccionada por su alta precisión y eficiente rendimiento. Esta herramienta permite transformar los caracteres de la imagen en texto digital y asegurar una base confiable para el siguiente proceso de conversión a voz.
- *Conversión de texto a voz.* Esta tecnología permite la generación artificial de voz a partir de texto escrito mediante sistemas computarizados (Taylor, 2009). En este módulo, el texto digitalizado obtenido del sistema OCR es procesado por el motor eSpeak, seleccionado por su eficiencia computacional, compatibilidad con dispositivos de asistencia y soporte multilingüe (eSpeak, s. f.). El resultado es una salida de audio clara y fluida que facilita el acceso auditivo al contenido textual procesado.

Integración del sistema

- *Integración del sistema mecánico.* Esta fase se centra en la construcción física del dispositivo, con el chasis y la carcasa previamente diseñados. En esta etapa, se ensamblan los componentes mecánicos de manera precisa para garantizar un ajuste adecuado de todas las piezas. El objetivo es lograr un sistema funcional y ergonómico que, al ser acoplado a las gafas, ofrezca comodidad y facilidad de uso, sin comprometer su desempeño.
- *Integración del sistema electrónico.* En esta fase, se procede a la conexión de todos los componentes electrónicos según el esquema de conexión diseñado en etapas anteriores. Esto incluye la interconexión de la Raspberry Pi, la cámara, los auriculares y la batería portátil, lo que asegura que cada uno de los elementos esté correctamente alimentado y funcionando en conjunto de manera eficiente.
- *Integración del software.* En esta etapa final, se instalará y configurará el *software* de visión por computadora en la Raspberry Pi, que incluye el

reconocimiento óptico de caracteres y la síntesis de texto a voz. Se realizarán pruebas exhaustivas para asegurar la correcta interacción del *software* con el *hardware*, a través del procesamiento de imágenes y la generación de la salida de audio.

Figura 5

Integración del dispositivo



RESULTADOS

Para evaluar el desempeño del *software* desarrollado, se realizaron pruebas con textos impresos en hojas de papel bond tamaño A4, bajo condiciones óptimas de iluminación. A diferencia del estudio realizado por Chinchero Iza (2019), en esta evaluación no se consideraron variaciones en las condiciones lumínicas.

El análisis se centró en cuatro factores clave que influyen en la eficiencia del reconocimiento óptico de caracteres:

1. *Tipos de textos*. Se evaluaron tres tipos diferentes de textos.
2. *Distancia de captura*. Se realizaron pruebas a 18 cm, 22 cm y 26 cm de distancia.
3. *Tamaño de la letra*. Se usaron tamaños de 10, 12 y 14 puntos.
4. *Tipos de fuente*. Se probaron las tipografías Arial, Times New Roman y Aptos.

En total, se llevaron a cabo 81 pruebas de acuerdo con todas las combinaciones de estos factores.

Textos de prueba

Para la evaluación, se utilizaron tres textos con distintas fuentes tipográficas. La Figura 6 muestra el texto de prueba 1 con la fuente Arial; la Figura 7, el texto de prueba 2 con Times New Roman; y la Figura 8, el texto de prueba 3 con Aptos.

Figura 6

Texto de prueba 1 con fuente Arial

La Unión Nacional de Ciegos del Perú (UNCP) es una organización sin fines de lucro que representa y defiende los derechos e intereses de las personas con discapacidad visual en el Perú. Su misión principal es promover la inclusión, igualdad de oportunidades y el desarrollo integral de las personas ciegas o con baja visión, fomentando su participación en todos los ámbitos de la sociedad.

La UNCP trabaja en colaboración con instituciones públicas, privadas y otras organizaciones de personas con discapacidad para lograr sus objetivos y mejorar la calidad de vida de sus asociados. Además, organiza talleres, eventos culturales y actividades recreativas para fomentar la integración social.

Figura 7

Texto de prueba 2 con fuente Times New Roman

La Universidad Nacional de Ingeniería (UNI) es una de las instituciones académicas más prestigiosas del Perú, reconocida por su excelencia en la formación de profesionales altamente calificados en las áreas de ingeniería, arquitectura y ciencias aplicadas. Fue fundada el 18 de marzo de 1876 por el ingeniero polaco Eduardo de Habich, inicialmente como la Escuela de Ingenieros de Construcción Civil y de Minas del Perú.

La Universidad Nacional de Ingeniería es un referente en la formación de profesionales y en la generación de conocimiento en el Perú, contribuyendo significativamente al progreso industrial, tecnológico y científico del país.

Figura 8

Texto de prueba 3 con fuente Aptos

La Facultad de Ingeniería Mecánica (FIM) de la Universidad Nacional de Ingeniería (UNI) es una de las unidades académicas más destacadas de la institución, especializada en la formación de ingenieros mecánicos y afines que contribuyen al desarrollo industrial, tecnológico y energético del Perú. Fundada como una de las primeras facultades de la UNI, la FIM tiene una larga tradición de excelencia académica y compromiso con la innovación tecnológica. Tiene como misión formar profesionales competentes con una sólida base científica y tecnológica, comprometidos con el desarrollo sostenible, la innovación y la solución de problemas industriales y energéticos del país.

Textos detectados

El reconocimiento OCR se evaluó en el entorno Thonny. Las figuras 9, 10 y 11 presentan el resultado del reconocimiento del texto de prueba 1 (fuente Arial, tamaño 10) capturado a 18 cm, 22 cm y 26 cm, respectivamente.

Figura 9

Reconocimiento OCR del texto de prueba 1 con fuente Arial, tamaño 10, capturado a 18 cm

Texto detectado: La Unión Nacional de Ciegos del Perú (UNCP) es una organización sin fines de lucro que representa y defiende los derechos e intereses de las personas con discapacidad visual en el Perú. Su misión principal es promover la inclusión, Igualdad de oportunidades y el desarrollo integral de las personas ciegas o con baja visión, fomentando su participación en todos los ámbitos de la sociedad.

La UNCP trabaja en colaboración con Instituciones públicas, privadas y otras organizaciones de personas con discapacidad para lograr sus objetivos y mejorar la calidad de vida de sus asociados. Además, organiza talleres, eventos culturales y actividades recreativas para fomentar la integración social.

Total de caracteres en el texto real: 694
Sustituciones (S): 3 -> [(232, 'i', 'I'), (427, 'i', 'I'), (462, 'o', 'a')]
Eliminaciones (D): 0 -> []
Inserciones (I): 0 -> []
Errores totales: 3
CER: 0.43%

Figura 10
Reconocimiento OCR del texto de prueba 1 con fuente Arial, tamaño 10, capturado a 22 cm

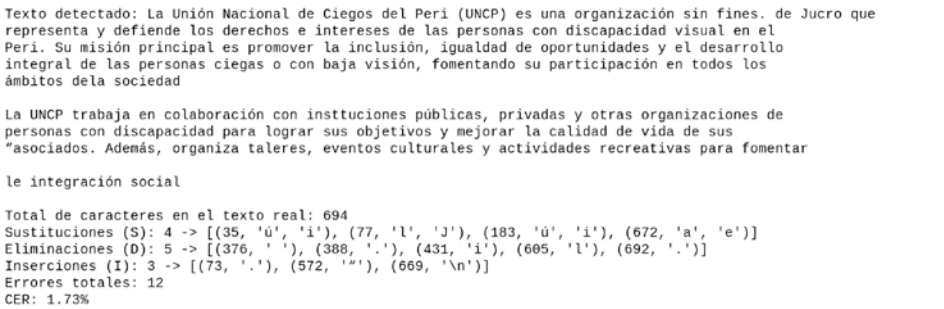
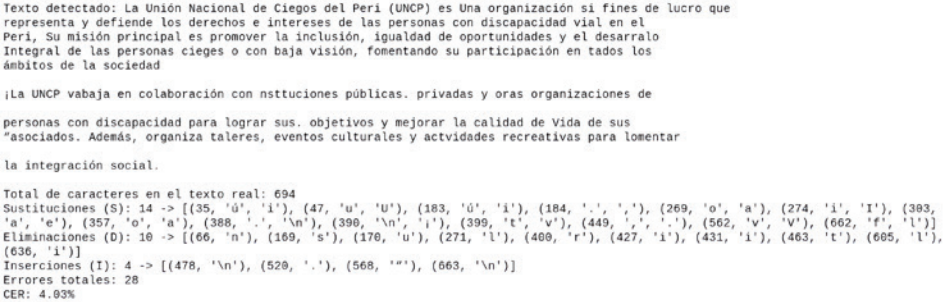


Figura 11
Reconocimiento OCR del texto de prueba 1 con fuente Arial, tamaño 10, capturado a 26 cm



Cálculo de la tasa de error por carácter

El cálculo de la tasa de error por carácter (CER) mide la proporción de caracteres erróneos con respecto al número total de caracteres del texto original. Se calcula mediante la siguiente ecuación (Vidal et al., 2023):

$$CER = \frac{S + D + I}{N} * 100$$

- Donde:
- S: número de sustituciones (caracteres incorrectamente reconocidos)
 - D: número de eliminaciones (caracteres omitidos)
 - I: número de inserciones (caracteres agregados)
 - N: número total de caracteres en el texto original

Para calcular el CER, se utilizó un proceso automatizado basado en la biblioteca Levenshtein en Python. El *software* compara el texto original con el texto detectado a través del método editops, el cual identifica las operaciones necesarias para transformar un texto en otro (Adjetey & Adu-Manu, 2021).

Las métricas de error se obtienen de la siguiente manera:

- *Sustituciones*. Se cuentan los caracteres en los que el OCR detectó un símbolo diferente al presente en el texto original.
- *Eliminaciones*. Se registran los caracteres que estaban en el texto original pero que el OCR no logró reconocer.
- *Inserciones*. Se contabilizan los caracteres que el OCR agregó de manera errónea y que no estaban en el texto original.

Este proceso es completamente automatizado, de modo que se elimina la necesidad de revisar manualmente cada palabra, lo que asegura precisión y eficiencia en la evaluación del desempeño del OCR.

Tabla 2
Tasa de error por carácter

Distancia	Tipo de letra	Tamaño	Texto 1		Texto 2		Texto 3	
			Caracteres	% CER	Caracteres	% CER	Caracteres	% CER
18	Arial	10	694	0,43	647	1,70	671	0,30
18	Arial	12	694	0,14	647	0,77	671	0,30
18	Arial	14	694	0,14	647	0,31	671	0,15
18	Times	10	694	0,58	647	2,01	671	1,19
18	Times	12	694	0,58	647	1,24	671	0,60
18	Times	14	694	0,43	647	0,77	671	0,30
18	Aptos	10	694	3,60	647	1,39	671	1,34
18	Aptos	12	694	0,58	647	0,31	671	0,30
18	Aptos	14	694	0,14	647	0,77	671	0,30
22	Arial	10	694	1,73	647	0,93	671	0,89
22	Arial	12	694	0,43	647	1,08	671	0,60
22	Arial	14	694	0,58	647	0,46	671	0,45
22	Times	10	694	3,17	647	4,02	671	2,53
22	Times	12	694	1,87	647	2,17	671	0,60
22	Times	14	694	0,54	647	0,93	671	0,45

(continúa)

(continuación)

Distancia	Tipo de letra	Tamaño	Texto 1		Texto 2		Texto 3	
			Caracteres	% CER	Caracteres	% CER	Caracteres	% CER
22	Aptos	10	694	3,03	647	9,58	671	2,53
22	Aptos	12	694	2,16	647	0,46	671	1,04
22	Aptos	14	694	0,58	647	1,39	671	0,75
26	Arial	10	694	4,03	647	6,66	671	4,02
26	Arial	12	694	1,01	647	1,70	671	1,04
26	Arial	14	694	2,45	647	0,93	671	0,75
26	Times	10	694	7,20	647	13,47	671	7,15
26	Times	12	694	3,31	647	11,76	671	2,09
26	Times	14	694	3,75	647	2,01	671	0,75
26	Aptos	10	694	7,35	647	15,15	671	6,56
26	Aptos	12	694	2,74	647	1,24	671	2,53
26	Aptos	14	694	1,01	647	1,70	671	1,19
			Promedio	1,98	Promedio	3,14	Promedio	1,51
			Promedio total					2,21

A partir de los datos de la Tabla 2, se calculó un CER promedio de 2,21 %. La precisión del sistema se obtiene mediante la fórmula: precisión = 1 - CER. Al sustituir el valor obtenido, se logra una precisión del 97,79 %, lo que indica un alto desempeño en el reconocimiento de caracteres.

DISCUSIÓN

En este proyecto se desarrolló un dispositivo de OCR y TTS capaz de convertir documentos impresos en audio, lo que facilitó el acceso a la información para personas con discapacidad visual. La evaluación del sistema se llevó a cabo con textos impresos en hojas de papel bond tamaño A4, bajo condiciones de iluminación óptimas.

Para analizar el desempeño del OCR, se consideraron cuatro factores clave: tipo de texto, distancia de captura (18 cm, 22 cm y 26 cm), tamaño de la letra (10, 12 y 14 puntos) y tipo de fuente (Arial, Times New Roman y Aptos). En total, se realizaron 81 pruebas, lo que permitió obtener una evaluación detallada del rendimiento del sistema en diferentes condiciones.

Comparación de tasa de error por carácter

El rendimiento del OCR varió significativamente según la tipografía utilizada, lo que afecta la precisión en el reconocimiento de caracteres. En las pruebas realizadas con Tesseract OCR, se identificaron las siguientes tasas de error:

- *Arial*. Registró la menor tasa de error con un 1,26 % en CER. Su diseño sin serifa y su estructura clara permitieron una segmentación más precisa de los caracteres, lo que redujo los errores en el reconocimiento. Según Malkadi et al. (2020), las fuentes sin serifa, como Arial, obtuvieron una mejor precisión en OCR en comparación con fuentes con serifa debido a la ausencia de detalles adicionales en los caracteres que podrían dificultar el reconocimiento.
- *Times New Roman*. Presentó la mayor tasa de error con un 2,8 % en CER. Esta fuente con serifa, aunque ampliamente utilizada en documentos formales, tuvo un 20 % menos de precisión en comparación con Arial en las pruebas de Tesseract, lo que indica que las serifas y detalles adicionales pueden interferir con la segmentación de caracteres (Malkadi et al., 2020).
- *Aptos*. Registró un 2,58 % en CER, lo que indica un desempeño inferior al de Arial. Sin embargo, el estudio de Malkadi et al. (2020) no profundizó en las razones específicas detrás del rendimiento de esta fuente y solo se limitó a señalar que la elección de la tipografía influye significativamente en la precisión del OCR.

En la Figura 5, se presenta el diseño del prototipo, cuya carcasa protege los componentes electrónicos. Este diseño no solo garantiza la integridad de los componentes, sino que también mejora la portabilidad del dispositivo.

En comparación con estudios previos, Luna Soria (2022) desarrolló un dispositivo con una carcasa estática y de grandes dimensiones, lo que restringía su portabilidad y dificultaba su uso para personas con discapacidad visual en su vida diaria. De manera similar, Arias Acosta (2022) diseñó un prototipo con una movilidad limitada, ya que sus dimensiones seguían siendo poco prácticas para un uso cotidiano.

En cambio, nuestro dispositivo fue desarrollado para ser compacto y de bajo costo, con dimensiones de 14 cm de largo, 10,2 cm de ancho y 8,5 cm de alto. Estas características lo hacen ligero y fácil de transportar, de modo que permiten su uso en diversos entornos sin comprometer la protección de los componentes electrónicos.

Durante las pruebas, el sistema TTS utilizó eSpeak, lo que resultó en una voz con un tono robótico. Este aspecto se puede optimizar en futuras versiones del dispositivo con la aplicación de técnicas avanzadas que mejoren la calidad de la conversión.

Limitaciones y futuras investigaciones

- *Calidad del TTS.* Actualmente, la voz generada por el motor de síntesis de texto a voz presenta un tono robótico, lo que afecta la experiencia del usuario. Para mejorar este aspecto, se recomienda explorar motores de TTS más avanzados, como Google Text-to-Speech, Amazon Polly o Microsoft Azure Speech, los cuales ofrecen voces más naturales y personalizables. Además, se podrían aplicar técnicas de procesamiento de señales para suavizar la entonación y mejorar la fluidez del habla generada.
- *Compatibilidad de la cámara.* Se presentaron problemas de compatibilidad entre la cámara Raspberry Pi V3 y el código ejecutado en Thonny, lo que llevó a optar por una webcam USB como alternativa. Debido a las limitaciones de rendimiento de la Raspberry Pi 3, probar otros entornos de desarrollo no fue viable, ya que la ejecución del código fue demasiado lenta. Para superar esta limitación en futuras versiones, se recomienda evaluar un *hardware* más potente, como la Raspberry Pi 5, que mejora la integración de la cámara sin comprometer el rendimiento del sistema.
- *Desarrollo de OCR con deep learning.* Como línea de investigación futura, se propone desarrollar un sistema OCR basado en *deep learning* según el enfoque de Namysl y Konya (2019). Esta solución elimina la necesidad de segmentación explícita de caracteres mediante el uso de redes neuronales profundas, las cuales aprovechan los datos sintéticos y técnicas de *data augmentation* para mejorar la robustez del reconocimiento en tipografías diversas y entornos complejos.
- *Reconocimiento de texto en entornos 3D complejos.* Futuras investigaciones podrían explorar la integración de estimación de profundidad y modelos de detección sin anclas para mejorar el OCR en entornos tridimensionales. Según Soans y Fukumizu (2024), la combinación de detección de objetos 3D y aprendizaje profundo optimiza la extracción de texto en condiciones de iluminación variable y ángulos complejos, lo que podría aplicarse a superficies no planas en estudios posteriores.

CONCLUSIONES

Se concluye que el modelado 3D es esencial para optimizar el diseño del prototipo y garantizar la compatibilidad de sus componentes antes de la fabricación mediante impresión 3D. Resulta fundamental que el diseño se ajuste con precisión a las dimensiones de la Raspberry Pi, lo que asegurará la sujeción adecuada de la batería Xiaomi Mi 3 Ultra Compact y el correcto ensamblaje de los botones en sus respectivos orificios. Estas mejoras estructurales permiten diseñar un dispositivo compacto, ligero y portátil, que optimiza su uso en diferentes entornos.

El algoritmo de OCR implementado mostró un desempeño eficiente en la conversión de texto a audio. No obstante, se identificó que su precisión depende en gran medida de condiciones óptimas de iluminación, lo que es un factor determinante para reconocer caracteres especiales con exactitud. Se detectó que ciertas letras, como la e y la o, pueden confundirse en algunos casos, lo que sugiere la necesidad de mejorar el algoritmo mediante técnicas avanzadas de procesamiento de imagen.

El prototipo desarrollado representa una herramienta de asistencia valiosa para personas con discapacidad visual, pues facilita su acceso a la información impresa de manera más autónoma. Las pruebas con usuarios invidentes confirmaron su utilidad, aunque señalaron la necesidad de optimizarlo para garantizar una experiencia completamente independiente. Mejoras futuras podrían incluir el uso de algoritmos de inteligencia artificial para fortalecer el reconocimiento de caracteres y la integración de un sistema de navegación guiada por voz para hacer el dispositivo aún más accesible y funcional en distintos escenarios.

En conclusión, este estudio contribuye al desarrollo de tecnologías accesibles al ofrecer un dispositivo compacto, funcional y de bajo costo, que podría beneficiar significativamente la independencia de personas con discapacidad visual.

CONFLICTOS DE INTERÉS

Los autores declaran no tener conflictos de interés.

CONTRIBUCIÓN DE LOS AUTORES

Cristhian Rafael Rosales: conceptualización, *data curation*, análisis formal, adquisición de fondos, investigación, *software*, validación, visualización, escritura-borrador original, redacción (revisión y edición). **José Salazar Minchan:** conceptualización, *data curation*, adquisición de fondos, metodología, *software*, escritura-borrador original. **Jorge Ortiz Porras:** administración de proyecto, recursos, supervisión, validación, redacción (revisión y edición).

REFERENCIAS

- Adjetey, C., & Adu-Manu, K. S. (2021). Content-based image retrieval using Tesseract OCR engine and levenshtein algorithm. *International Journal of Advanced Computer Science and Applications*, 12(7), 666-675. <https://dx.doi.org/10.14569/IJACSA.2021.0120776>
- AlSaid, H., AlKhatib, L., AlOraih, A., AlHaidar, S., & Bashar, A. (2019). *Deep learning assisted smart glasses as educational aid for visually challenged students*

[Presentación de escrito]. 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordania. <http://dx.doi.org/10.1109/ICTCS.2019.8923044>

Arias Acosta, M. D. (2022). *Diseño e implementación de gafas inteligentes como herramienta de asistencia para lectura y reconocimiento del entorno para personas con discapacidad visual utilizando inteligencia artificial* [Tesis de licenciatura, Universidad de las Fuerzas Armadas]. Repositorio institucional de la Universidad de las Fuerzas Armadas de Ecuador. <http://repositorio.espe.edu.ec/handle/21000/35773>

Basantes Varela, D. A., & Chalaco Chamba, E. A. (2019). *Desarrollo de un prototipo de gafas para lectura de texto con visión artificial que asista a personas con discapacidad visual* [Tesis de licenciatura, Universidad Politécnica Salesiana]. Repositorio Institucional de la Universidad Politécnica Salesiana. <http://dspace.upse.edu.ec/handle/123456789/17841>

Chinchero Iza, J. L. (2019). *Desarrollo de un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes* [Tesis de licenciatura, Universidad Tecnológica Israel]. Repositorio Digital de la Universidad de Israel. <http://repositorio.uisrael.edu.ec/handle/47000/2117>

Consejo Nacional para la Integración de la Persona con Discapacidad. (2025). *Registro Nacional de la Persona con Discapacidad*. Observatorio Nacional de la Discapacidad. <https://observatorio.conadisperu.gob.pe/>

Cuesta, I. I., Martínez-Pañeda, E., Díaz, A., & Alegre, J. M. (2019). The essential work of fracture parameters for 3D printed polymer sheets. *Materials & Design*, 181, 107968. <https://doi.org/10.1016/j.matdes.2019.107968>

eSpeak. (s. f.). *eSpeak text-to-speech software*. <http://espeak.sourceforge.net/>

Howse, J., & Minichino, J. (2020). *Learning OpenCV 4 Computer Vision with Python 3. Get to grips with tools, techniques, and algorithms for computer vision and machine learning* (3.^a ed.). Packt Publishing.

Instituto Nacional de Estadística e Informática. (2019). *Perfil sociodemográfico de la población con discapacidad, 2017*. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1675/libro.pdf

Karmel, A., Sharma, A., Pandya, M., & Garg, D. (2019). IoT based assistive device for deaf, dumb and blind people. *Procedia Computer Science*, 165, 259-269. <https://doi.org/10.1016/j.procs.2020.01.080>

- Kuriakose, B., Shrestha, R., & Sandnes, F. E. (2022). Tools and technologies for blind and visually impaired navigation support: A review. *IETE Technical Review*, 39(1), 3-18. <https://doi.org/10.1080/02564602.2020.1819893>
- López Jiménez, S. (2019). *Ayudas electrónicas para pacientes con discapacidad visual* [Tesis de maestría, Universidad de Valladolid]. Repositorio de la Universidad de Valladolid. <http://uvadoc.uva.es/handle/10324/37124>
- Luna Soria, D. O. (2022). *Desarrollo de un prototipo de dispositivo para reconocimiento de texto tipo imprenta y su conversión en audio orientado a personas con discapacidad visual mediante técnicas de visión artificial* [Tesis de licenciatura, Escuela Superior Politécnica de Chimborazo]. Repositorio Institucional de la Escuela Superior Politécnica de Chimborazo. <http://dspace.esPOCH.edu.ec/handle/123456789/21272>
- Malkadi, A., Alahmadi, M., & Haiduc, S. (2020, June). A study on the accuracy of OCR engines for source code transcription from programming screencasts [Presentación de escrito]. Proceedings of the 17th International Conference on Mining Software Repositories, Nueva York, Estados Unidos. <https://doi.org/10.1145/3379597.3387468>
- Manage, P., Ambe, V., Gokhale, P., Patil, V., Kulkarni, R. M., & Kalburgimath, P. R. (2020, 3-5 de diciembre). *An intelligent text reader based on python* [Presentación de escrito]. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India. <https://doi.org/10.1109/ICISS49785.2020.9315996>
- Namysl, M., & Konya, I. (2019, 20-25 de septiembre). *Efficient, lexicon-free OCR using deep learning* [Presentación de escrito]. 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia. <https://doi.org/10.1109/ICDAR.2019.00055>
- Navarro Pérez, V. M. (2021). *Sistema portátil de reconocimiento de texto para la asistencia de personas ciegas o con dificultad en la visión* [Tesis de grado, Universidad Rey Juan Carlos]. Repositorio de la Universidad Rey Juan Carlos. <http://hdl.handle.net/10115/18188>
- Organización Mundial de la Salud. (2023). *Ceguera y discapacidad visual*. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Raspberry Pi Foundation. (2023). *Raspberry Pi 3 Model B+. Product brief*. <https://datasheets.raspberrypi.com/rpi3/raspberry-pi-3-b-plus-product-brief.pdf>
- Raspberry Pi Foundation. (2024). *Raspberry Pi Camera Module 3. Product brief*. <https://datasheets.raspberrypi.com/camera/camera-module-3-product-brief.pdf>

- Smith, R. (2007, septiembre). *An overview of the Tesseract OCR engine* [Presentación de escrito]. 9th International Conference on Document Analysis and Recognition (ICDAR), Curitiba, Brasil. <https://doi.org/10.1109/ICDAR.2007.4376991>
- Soans, R., & Fukumizu, Y. (2024). Custom anchorless object detection model for 3D synthetic traffic sign board dataset with depth estimation and text character extraction. *Applied Sciences*, 14(14), 6352. <https://doi.org/10.3390/app14146352>
- Taylor, P. (2009). *Text-to-speech synthesis*. Cambridge University Press.
- Vidal, E., Toselli, A. H., Ríos-Vila, A., & Calvo-Zaragoza, J. (2023). End-to-end page-level assessment of handwritten text recognition. *Pattern Recognition*, 142, 109695. <https://doi.org/10.1016/j.patcog.2023.109695>
- Wong Cam, C. F., & Wong Morales, C. A. (Eds.). (2019). *Enfermedades de los ojos y ceguera en el Perú*. American Offset Editores. <https://wongoftalmologos.com.pe/wp-content/uploads/2024/04/Libro-Enfermedad-y-Ceguera-en-el-Peru-1.pdf>
- Xiaomi. (s. f.). *Manual de usuario de 10 000 mAh Mi 18W Fast Charge Power Bank 3*. https://i01.appmifile.com/webfile/globalimg/Global_UG/Mi_Ecosystem/10000mAh_Mi_18W_Fast_Charge_Power_Bank_3/es-ES_V1.pdf
- Yataco Marín, R. M. (2022). Tiflotecnología y el acceso a la información de las personas con discapacidad visual. *Fénix*, (50), 76-90. <https://doi.org/10.51433/fenix-bnp.2022.n50.p76-90>