

aSISTA. Aplicación de Asistencia a Personas con Discapacidades Visuales en Dispositivos Móviles

aSISTA. An Assistance Mobile Application for Visually Impaired People

Maikel Salas Zaldívar
La Habana, Cuba
maikelsz8143@nauta.cu

Recibido: 21.03.2023 | Aceptado: 30.05.2023

Palabras Clave

aplicación
asistencia
discapacitado
inteligencia artificial
dispositivo móvil
reconocimiento

Resumen

La vista es uno de los sentidos de mayor importancia para el ser humano por lo que las personas con discapacidad visual afrontan dificultades para su desenvolvimiento día a día. Sin embargo, muchas de estas personas poseen dispositivos móviles, que, si bien su forma de interacción no está explícitamente diseñada para estas ser utilizado por estas personas, logran interactuar con ellos gracias a tecnologías de accesibilidad como lectores de pantalla y sintetizadores de voz. Estos dispositivos en su mayoría poseen una cámara digital que permite capturar imagen y video, así como otros sensores. Haciendo uso de estas capacidades y mediante la aplicación de técnicas de procesamiento de imagen y reconocimiento de patrones, se ha desarrollado una aplicación móvil, que funcionando en móviles y tabletas de la plataforma Android, les permite a las personas con discapacidades visuales reconocer billetes, tarjetas bancarias, objetos, reconocer el entorno y navegarlo, así como colores y niveles de iluminación. Toda la interacción con la aplicación se realiza mediante toques y deslizamientos y la realimentación es mediante voz sintetizada, sonidos o reacciones táctiles. De esta manera, estas personas ganan en independencia y mejora su inserción social...

Keywords

application
assistance
visual impairment
artificial intelligence
mobile device
recognition

Abstract

Sight is one of the most important senses for human beings and for this reason, people with visual impairments face difficulties for their day-to-day development. However, many of these people have mobile devices, which, although their form of interaction is not explicitly designed to be used by them, they manage to interact with them thanks to accessibility technologies such as screen readers and voice synthesizers. These devices mostly have a digital camera that allows image and video capture, as well as other sensors. Making use of these capacities and through the application of image, pattern recognition and artificial intelligence techniques, a mobile application has been developed, which, working on Android mobile phones and tablets, allows people with visual disabilities to recognize bills, cards banking, objects, recognizing the environment and navigating it, as well as colors and lighting levels. All interaction with the app is done by tapping and swiping and feedback is through synthesized voice, sounds, or tactile reactions. In this way, these people gain independence and improve their social integration.

1. Introducción

La vista es un sentido de mucha importancia para el ser humano que nos permite desenvolvernó con mayor facilidad en las actividades diarias. Según estimaciones de la Organización Mundial de la Salud (OMS), en el mundo había aproximadamente 285 millones de personas (más del 4% de la población mundial) con alguna discapacidad visual, de las cuales 39 millones eran ciegas (WHO, 2012) (WHO, 2019). Además, reportaba que el 82% de las personas que padecían ceguera tenían 50 o más años de edad (Bourne, 2017) lo que hace aún más difícil su desenvolvimiento social. En el caso de Cuba, hay más de 30 mil personas ciegas o con discapacidad visual asociadas a la ANCI (Asociación Nacional del Ciego) (Ecured, 2021).

El creciente número de teléfonos móviles los han convertido en el principal medio de comunicación y fuente de información. Sin embargo, los usuarios con discapacidad visual no pueden utilizar todas las aplicaciones que se desarrollan. A diferencia de los usuarios videntes enfrentan una obvia dificultad para interactuar con teléfonos móviles de pantalla táctil. La inaccesibilidad de los teléfonos móviles se debe principalmente a sus técnicas de interacción, que requieren que el usuario ubique visualmente los objetos en la pantalla (Rodríguez, 2020).

Para garantizar que las aplicaciones de software y los sitios web sean accesibles para las personas con discapacidad visual, existen una serie de pautas de accesibilidad que los desarrolladores deben seguir. Por ejemplo, las Pautas de accesibilidad al contenido web (W3C, 2018) brindan recomendaciones detalladas para hacer que los sitios web sean accesibles para los usuarios con discapacidades visuales. Estas pautas cubren todo, desde el contraste de color hasta la navegación del teclado.

La “Accesibilidad Móvil” se refiere a hacer sitios web y aplicaciones más accesibles a personas con discapacidades cuando usan teléfonos inteligentes u otros dispositivos móviles. Se tienen en cuenta posibles problemas como pantallas táctiles, pequeñas, diferentes formas de interacción, así como comportamiento en diversos entornos (W3C, 2023).

Los usuarios discapacitados visuales han tenido que crear métodos propios para utilizar teléfonos móviles o encontrar alternativas. Este problema puede resolverse con tecnologías de asistencia que permiten a estos usuarios acceder a los teléfonos móviles más fácilmente (Sagale, 2018). En el caso de los teléfonos inteligentes, por ejemplo, utilidades para lectura de pantalla y síntesis de voz pueden ser descargadas desde las tiendas de aplicaciones móviles. Estas utilidades permiten la interacción con el dispositivo entregando respuestas sonoras (palabras, sonidos) o táctiles / aptica (vibraciones). Así los usuarios ciegos pueden conocer qué ocurre en la pantalla y con qué interactúan, y de esta manera utilizar muchas más

aplicaciones móviles. También pueden emplear la cámara digital para capturar imágenes utilizadas en muchas tareas de reconocimiento.

Las personas discapacitadas visuales deben afrontar diariamente diferentes situaciones que generan tensión física y psicológica condicionando su calidad de vida (Augustin, 2007). Una de estas situaciones es reconocer y diferenciar correctamente la denominación de los billetes. Las monedas metálicas se pueden reconocer potencialmente en función de la forma, pero los billetes en la mayoría de las monedas del mundo tienen un tamaño similar y muchas veces también colores similares. Esto puede causar problemas incluso a las personas no totalmente ciegas. En estos casos también se puede enmarcar el reconocimiento de tarjetas bancarias, de pago u otros tipos.

La orientación y comprensión del entorno más allá del alcance de su baja visión, del bastón asistente o de la capacidad auditiva es una dificultad para los invidentes. Otra dificultad está en los obstáculos que encuentran durante la marcha, con los que pueden tropezar e incluso provocarles accidentes. Otras dificultades son el conocer el color de un objeto o el nivel de iluminación local.

Hay muchos tipos de tecnología de asistencia disponibles para ayudar a las personas con discapacidad visual a interactuar con las computadoras. Por ejemplo, los lectores de pantalla, que son programas de software que leen en voz alta el texto de la pantalla, pueden ayudar a los usuarios con discapacidades visuales a navegar a través de sitios web y aplicaciones de software. También hay pantallas braille, que utilizan una serie de pines para mostrar texto en braille, así como software de reconocimiento de voz y otras herramientas.

Existen algunas aplicaciones especializadas de asistencia a las personas con discapacidad visual que hacen uso de funciones avanzadas de procesamiento de imagen, reconocimiento de patrones e inteligencia artificial. Algunas están disponibles en tiendas de software como Google Play (para Android) o App Store (para iOS), otras en la web propia del desarrollador, otras son solo resultado de desarrollos experimentales. Algunos ejemplos son:

Currency Identifier (IDEAL, 2023), Money Reader (LookTel, 2023), ViaOptaDaily (), Blind-Droid Wallet (Mirwebsistem, 2020), TapTapSee (CloudSight, 2023), NavEye (Alabri, 2014), NavCog (Sato, 2017), Smart Backpack (Cruz, 2018), Sightless Helper (Hossain, 2020), Lozokout (Google, 2023), Seeing AI (Microsoft, 2023).

Algunas de estas aplicaciones van destinadas a reconocer papel moneda, navegar el entorno, reconocer textos, conectar con personas asistentes voluntarios, entre otras.

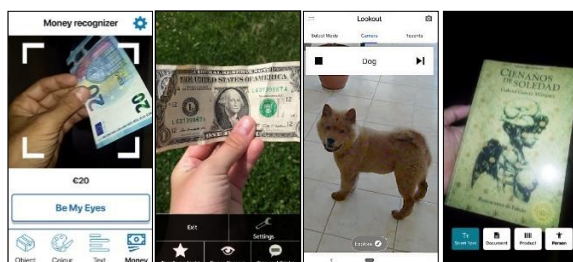


Figura 1: Algunas aplicaciones de asistencia. Se observan controles y visualización innecesarios para sus usuarios típicos.

Usualmente implementan una única función, por ejemplo, reconocer billetes. Pocas tienen alguna adicional. Estas funciones disjuntas pueden obligar a los usuarios a tener que disponer de dos o más aplicaciones si quieren cubrir la mayor cantidad de necesidades de asistencia posibles. Esto implica para ellos tener que pasar de unas aplicaciones a otras, las que pueden tener diferentes maneras de funcionar, interactuar o realimentar información.

Lookout de Google y *SeeingAI* de Microsoft son las más completas y avanzadas al proveer varias funcionalidades, aunque en nuestra opinión, también adolecen de varias limitantes.

- *SeeingAI* solo funciona en el sistema operativo iOS. En el caso de Cuba, este tiene poca presencia, <10%.
- Poseen interfaces de usuario visuales, innecesarias o inconvenientes para los usuarios objetivo.
- En el caso de reconocer billetes no incluyen la moneda cubana, y por supuesto, tampoco las tarjetas.
- Algunas funciones requieren conectividad servicios de terceros, con implicaciones de posibles intermitencias (por ejemplo, por no acceso a la red de datos), de costes (tráfico de datos) y de privacidad.

La mayoría de estas aplicaciones no referencian la base científico-técnica que emplean.

Esto llevó a plantearse el desarrollo de una aplicación unitaria que, ejecutándose en un dispositivo móvil, cubriera la mayor cantidad de posibles necesidades de asistencia de las personas con discapacidades visuales, fuese de fácil y amplio despliegue, cubriera además algunas necesidades específicas al contexto nacional y tuviera en cuenta los aspectos mejorables de otras aplicaciones existentes.

2. Desarrollo

Previamente, se desarrolló una aplicación para dispositivos móviles para hacer accesible a las personas con discapacidades visuales el reconocimiento automático de billetes cubanos, de euro, dólar y tarjetas de diferentes tipos usadas en Cuba (y par de funcionalidades adicionales) empleando los procedimientos antes descritos. Esta se denomina qBillete (Salas, 2021) (Salas, 2022) y está destinada a funcionar sobre el sistema operativo Android. Se analizaron las bases de este tipo de aplicaciones

(Rashid, 2013) (Solymar, 2011) definiéndose los requerimientos básicos, así como las facilidades adicionales que enriquecieran e hicieran más funcional la aplicación. Se tuvieron en cuenta por igual las funcionalidades de algunas de las aplicaciones existentes.

La aplicación original fue ampliada incluyendo nuevas funcionalidades y capacidades. Además del reconocimiento de billetes y tarjetas se incluyó reconocimiento de objetos, su posición y distancia estimada, la descripción general del entorno, la navegación por etiquetas descriptivas, el reconocimiento de textos, la transcripción de dictados, el reconocimiento de colores y de niveles de iluminación. Las formas de interacción se ampliaron, incluyéndose comandos de voz.

Varios requerimientos deben cumplirse en una aplicación móvil de esta índole (Dobosz, 2017):

- Una aplicación móvil destinada a personas con discapacidad visual requiere de una interactividad sencilla, que la navegación por sus funcionalidades y opciones sea de manera simple y directa.
- Un usuario espera que una aplicación de asistencia sea precisa, y genere poca o ninguna incertidumbre en sus operaciones y respuestas.
- Este tipo de aplicación posee necesidades de inmediatez. Por ejemplo, el caso de pagos inmediatos requiere de rápidas verificaciones de los billetes.
- Teniendo en cuenta las limitantes de los usuarios de estas aplicaciones, se requiere una realimentación de las acciones realizadas, las operaciones ejecutadas por la aplicación y las respuestas que esta entrega, así como ayudas, de maneras sonoras (voz sintética o grabada, pitidos), y apticas (vibración).

2.1 Desarrollo para Android

Android es un sistema operativo lanzado en 2007, basado en una versión modificada del núcleo del sistema operativo Linux y otros softwares de código abierto. Está diseñado principalmente para dispositivos móviles con pantalla táctil, como teléfonos inteligentes y tabletas. Es desarrollado por un consorcio conocido como Open Handset Alliance (OHA, 2023) y con el respaldo comercial de Google. Según datos recabados en internet, a marzo de 2023 a nivel mundial Android alcanzaba una cuota de mercado de un 72% (Statcounter, 2023a). En el caso de Cuba este valor alcanzaba 91% (Statcounter, 2023b). Según datos directos mostrados por ETECSA, el operador nacional de telecomunicaciones, nueve de los diez fabricantes de móviles más usados en la red móvil del país usaban Android como sistema operativo (Cubadebate, 2020). Esto está dado fundamentalmente por el menor coste de propiedad, facilidades de uso e intercambio de información, y en el caso de Cuba además a la existencia de aplicaciones específicas del contexto nacional solo existentes para esta plataforma, además de facilidades de modificación y

adaptación a necesidades locales y las mayores posibilidades de reparación a un menor costo.

Como entorno integrado de desarrollo se empleó Android Studio (Google, 2023a), con el código funcional escrito en lenguaje de programación Java.

Para el procesamiento de imagen y reconocimiento de patrones se emplea la biblioteca de código abierto OpenCV (OpenCV, 2023) compilada para Android y usando su API Java. Emplear OpenCV permite tener disponible en una única biblioteca funciones optimizadas para la mayoría de las operaciones primarias de procesamiento de imagen e inteligencia artificial a realizar.

2.2 Diseño

El diseño centrado en el usuario (Gould, 1985) es un componente crítico de la interacción humano-computadora para personas con discapacidades visuales. Implica diseñar interfaces de usuario que sean intuitivas y fáciles de usar, y que tengan en cuenta las necesidades y preferencias de este tipo de usuarios (Leo, 2016). Esto podría incluir características como texto grande y de alto contraste, menús y navegación bien organizados y comentarios de audio claros (Qureshi, 2019).

Mediante una recopilación de información al respecto presente en otros trabajos se definieron las bases comunes del diseño, las consideradas innecesarias o redundantes así como algunas nuevas.

La interfaz gráfica de la aplicación es mínima, de una simplificación máxima. No hay botones, ni letreros ni visualización innecesaria (algo que consume energía). Si bien por las características de sus usuarios la pantalla se puede mantener en negro, aquí muestra la simbología de la función activa, más bien por estética.

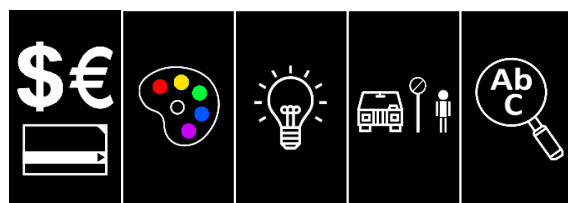


Figura 2: Simbología en pantalla de algunas de las funciones.

Al reconocer billetes o tarjetas se muestra el último resultado en caracteres de alto contraste. Al reconocer colores, la pantalla se ilumina intensamente con el color reconocido. Al realizar dictados, se muestra el texto transcrito.



Figura 3: Interfaz de usuario mostrando el resultado de una operación de reconocimiento de billetes.

2.3 Funcionalidades desarrolladas

La aplicación es capaz de realizar el reconocimiento de varias monedas y sus denominaciones, de diferentes tarjetas magnéticas, de diferentes colores, niveles de iluminación, objetos y su distancia estimada, textos/palabras, y transcribir dictados.

En el desarrollo de estas funcionalidades se han empleado técnicas de procesamiento de imagen, de reconocimiento de patrones, estadísticas, y de inteligencia artificial basadas en redes neuronales.

2.3.1 Reconocer billetes y tarjetas

Para el reconocimiento de billetes y tarjetas las imágenes se capturan continuamente, y se estima su calidad. Si la imagen no tiene calidad suficiente por razones tales como desenfoco o posición se alerta al usuario, la imagen se descarta y se pasa a capturar otra. De las imágenes aceptadas se extraen las características identificativas más robustas y se comparan con las plantillas almacenadas. La denominación con mayor correlación se toma como la correcta y se le anuncia al usuario, además de representar el número en caracteres grandes y de color contrastado.

2.3.2 Reconocer colores

Para el reconocimiento de color se toma el 75% central de la imagen captada. Esto es para reducir la influencia de partes de la imagen que no esté centrada en los objetos o regiones de interés. El color en las imágenes captadas por la cámara puede sufrir corrimientos debido a problemas como el color de la iluminación, la calidad de la cámara y como esta reacciona al espectro luminoso. Por ello se aplican hasta dos correcciones automáticas de color o balance de blanco: el propio del dispositivo si dispone de ello, y otro por software. Estas correcciones reducen las desviaciones del color real de la escena y así los errores en el reconocimiento. Otro factor que afecta la precisión es la especularidad de la superficie, que

pueden hacer parecer las superficies como grises o blancas a la cámara.

Creando un histograma de tono/saturación de los colores, se localiza en la imagen el color con mayor presencia, es decir, el más abundante y se compara con unos rangos preestablecidos para determinar los siguientes colores: rojo, naranja, amarillo, verde, azul, púrpura, blanco, gris y negro.

2.3.3 Reconocer niveles de iluminación

Para el reconocimiento del nivel de iluminación se emplea el sensor especializado incluido en la mayoría de los móviles moderno. Este sensor se encuentra usualmente en la parte frontal de los dispositivos móviles y tiende a ser bastante direccional, es decir, responde más a la luz que le llega directamente en un ángulo estrecho.

Los valores del nivel de iluminación se entregan en Lux (NIST, 2023), una unidad derivada del Sistema Internacional de Unidades que define el flujo luminoso por unidad de área. Con una calibración experimental se fijaron varios rangos de niveles de iluminación: desde oscuro (4 lux o menos), muy baja, baja, media, alta, a muy alta (1000 lux o más). El valor de la iluminación captada se compara con estos rangos y se entrega el resultado correspondiente mediante la voz sintetizada.

2.3.4 Reconocer objetos y describir escena

El reconocimiento de objetos está basado en un modelo red neuronal muy optimizado, del tipo YOLO (dogquiqui, 2021). Permite detectar entidades de unas 80 clases. Para evitar confusiones y ambigüedades, de estas clases se eliminaron las menos probables a ser encontradas en la vida cotidiana.

La imagen captada por la cámara es pasada a la red neuronal, que devuelve un conjunto de regiones, sus clases y un nivel de confianza.

Las detecciones que superen un umbral de confianza pasan a una estimación de la distancia a que se encuentra la entidad detectada. Esto se realiza mediante operaciones trigonométricas de odometría visual que emplean parámetros intrínsecos del sistema óptico del dispositivo como la distancia focal y una “altura media” del objeto. Se eligen hasta tres detecciones prioritarias más cercanas, con distancia estimada inferior a 5 metros, de las que se localiza entonces su posición, dígame a la izquierda, al frente o a la derecha. También pueden localizarse hacia arriba o abajo, pero esto se ha limitado para no abrumar al usuario y ganar tiempo.

Finalmente, con un orden de prioridad, la aplicación dice lo detectado, la distancia estimada y la posición relativa. Por ejemplo, tomando el caso mostrado en la figura 4, la

aplicación diría: “*persona, derecha, dos, persona, frente, cuatro, auto, derecha, cinco*”

La alerta sobre los objetos cercanos, su posición y distancia pueden permitir al usuario navegar su entorno con algo más de seguridad. Por ejemplo, disponiendo el dispositivo hacia el frente puede caminar mientras le va diciendo los posibles obstáculos a su paso. Con ayuda del sensor de aceleración disponible en algunos dispositivos (ver más adelante) puede incluso detectarse peligro de colisión y emitirse una alerta.

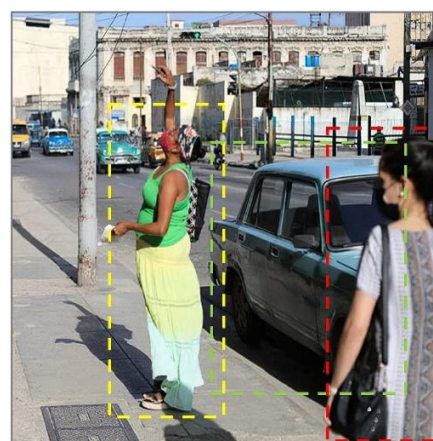


Figura 4: Detección en escena. Esto no se muestra en pantalla

Para una mejor escucha en entornos ruidosos como lugares muy transitados, pueden emplearse audífonos, preferiblemente del tipo de transducción ósea que no atenúan o impiden escuchar los sonidos de los alrededores, los cuáles son útiles a las personas ciegas para orientarse.

La “descripción de escena” se realiza a partir de la detección de objetos anterior, pero sin límite en cuanto a la cantidad o distancia máxima.

Con las clases determinadas para todas las detecciones válidas se elabora una sentencia descriptiva, que tiene en cuenta el prioridades, número y género de las detecciones. Un número inferior o igual a tres detecciones de una misma clase se anuncia en su cantidad específica. Un número mayor que tres, se anuncia como “varios” o “varias” según el género. De igual manera ocurre con una sola detección de una clase, se construye como “un/uno/una/unos/unas” según sea el caso. La distancia y la posición no se tienen en cuenta en este caso.

Tomando de nuevo la figura 4, en este caso la aplicación describiría la escena como: “*dos personas y varios autos*”.

2.3.5 Reconocer texto

La aplicación es capaz de detectar textos de tipografía variada dispersos en el entorno y reconocerlo. Las regiones resultado del detector de textos que cumplan cierto nivel mínimo de

certidumbre, pasan al reconocimiento como tal del texto mediante la biblioteca Tesseract (Ooms, 2023)



Figura 5: Detección y reconocimiento de texto.

Para evitar confusiones o abrumar al usuario con una larga narración, actualmente se limita el reconocimiento hasta unas 6 palabras de una longitud entre 3 y 14 caracteres. Estas son dichas por el sintetizador de voz.

2.3.6 Transcribir dictado

La función de transcripción de dictados, tal como la de comandos de voz (ver más adelante), emplea el *toolkit* VOSK (Nickolay, 2020). Se hace un reconocimiento de voz y convierte lo dicho en un texto, hasta cierta longitud. Esto puede ser especialmente útil para las personas que tienen dificultades para escribir en un teclado de pantalla táctil.

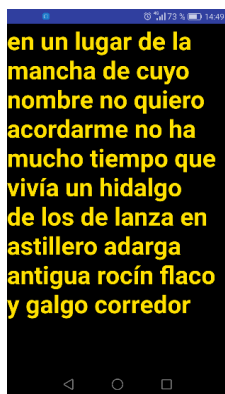


Figura 6: Ejemplo de la transcripción de un dictado.

Mientras el usuario mantiene activa la función y habla de manera clara, el texto transcrito va apareciendo en pantalla. Al finalizar de hablar, detener la operación o el texto llegar al límite establecido, el texto se copia a memoria y se repite por la voz sintetizada para que el usuario pueda verificar que sea correcto. El texto copiado puede ser luego empleado en mensajes de texto, correos o al mantenerse mostrado en pantalla, ser incluso una vía de comunicación de una persona discapacitada visual con una persona sorda.

2.3.7 Uso de los sensores

Los dispositivos móviles modernos poseen una serie de sensores que entregan datos para ser usados por las aplicaciones. Algunos de estos sensores son los de niveles de iluminación, acelerómetros, giroscopios, compás magnético, posicionamiento global y otros. No todos los sensores son incluidos en todos los dispositivos.

La información proporcionada por estos sensores puede emplearse para realizar análisis de estados o correcciones de las entradas de datos a las funcionalidades. Por ejemplo, las funciones de reconocimiento de objetos/escena, textos y etiquetas, requieren que la imagen mantenga una orientación vertical. Usando el giroscopio o el acelerómetro (que mide también la aceleración gravitatoria) se establece el vector vertical del dispositivo, lo que permite rotar la imagen anulando en lo posible la inclinación que introduce el usuario al manipularlo. El acelerómetro también permite saber cuándo el usuario está caminando y estimar su velocidad y junto a la detección de objetos estimar posibles colisiones y emitir alertas. También se emplea el sensor de iluminación, como se explicó anteriormente en otro punto. Estos y otros sensores pueden emplearse de diversas maneras para la asistencia, algunas están planeadas y se describen más adelante.

2.4 Privacidad

En los pasados años y actualmente, las consideraciones de seguridad y privacidad en las aplicaciones informáticas de todo tipo han tomado más relevancia que nunca. Los datos como imágenes, sonidos, localización, etc, enviados a sistemas remotos de terceros, siendo procesados e incluso almacenados a veces indefinidamente por servicios fuera del control del usuario, plantean muchas problemáticas de privacidad. Varios casos negativos relevantes y otros muchos algo menos han ocurrido en este contexto, desde accesos indebidos a los datos a explotación de los mismos por entes no autorizados, pasando por la venta de estos datos por parte de sus supuestos guardianes.

Otro problema del procesamiento “*on-the-cloud*” es la necesidad incremental de capacidad de procesamiento en los centros de datos, dígame necesidad de servidores, hardware especializado, aceleradores de procesos, y muy importante, los requerimientos energéticos para su funcionamiento. Todo esto lleva a incrementos en los costes de despliegue y operación.

Por otra parte, las capacidades “*on-the-edge*” se han incrementado. Los dispositivos, incluidos los móviles, han ido ganando cada vez más en capacidades de cómputo, memoria y almacenamiento, incluso incorporando unidades especializadas en la aceleración eficiente de diversas funciones avanzadas de procesamiento de imagen/video e inteligencia artificial.

Esto ha conducido a que muchas aplicaciones y servicios de requerimientos contenidos se hayan mantenido funcionando en los dispositivos, o incluso hayan regresado a ellos total o parcialmente desde previos despliegues en “en la nube” (Suliman, 2022).

Teniendo todo esto en cuenta, la aplicación desarrollada no realiza actualmente ningún acceso a datos ni servicios externos. Todo el software, procesamiento, datos y modelos están contenido y se realiza dentro de la aplicación y el dispositivo para su uso local. Las imágenes captadas para ser procesadas y analizadas se mantienen dentro del dispositivo. De hecho, la aplicación no almacena imágenes, las captura con la cámara y directamente las procesa.

El aspecto negativo es que la aplicación requiere entonces contener todas las bibliotecas, modelos y recursos adicionales necesarios para las funcionalidades incorporadas, lo que implica un aumento del tamaño de la aplicación. También implica consumo energético adicional y generación de calor durante la ejecución de las funcionalidades.

3. Interacción

La interacción humano-computadora estudia cómo los humanos interactúan con las computadoras y otras tecnologías digitales, combinando investigaciones y teorías de la informática, el diseño gráfico, la psicología, la ergonomía, la ingeniería y otros campos relacionados para diseñar y crear sistemas informáticos que sean fáciles e intuitivos de usar y que permitan a los usuarios realizar sus tareas de manera eficiente y eficaz. Esto implica al diseño de las interfaces de usuario, el diseño de la experiencia del usuario, las pruebas de usabilidad, los sistemas de entrada/salida y el software y el hardware que los respaldan (Rogers, 2023). Implica diseñar y evaluar sistemas informáticos para garantizar que sean accesibles, utilizables y que satisfagan las necesidades de los usuarios.

El desarrollo de interfaces que sean accesibles y fáciles de usar para las personas con discapacidad visual incluye el uso de fuentes grandes, esquemas de color de alto contraste y formas alternativas de entrada, como atajos de teclado o comandos de voz, así como formas de realimentación que puedan asimilar más fácilmente como las sonoras y las táctiles. Se tiene en cuenta cómo los usuarios con discapacidad visual interactúan con un sistema y se centran en sus necesidades específicas para garantizar que los sistemas que diseñan sean accesibles y fáciles de usar. Esto implica realizar pruebas de usuario con participantes con discapacidad visual para identificar formas de mejorar el diseño de un sistema, como proporcionar mejores comentarios o utilizar métodos de entrada alternativos.

El usuario emplea la aplicación interactuando por tacto y recibiendo realimentación sonora de los resultados y operaciones.

3.1 Interacción manual

La mayor parte de la interacción con la aplicación y sus funcionalidades es de manera manual, siguiendo en lo posible las formas de manipulación a que están acostumbrados los usuarios objetivo al emplear otras aplicaciones de asistencia.

Al iniciar la aplicación se emite un mensaje de bienvenida y anuncia la configuración activa. El usuario puede entonces realizar deslizamientos en pantalla hacia los lados para acceder a las funciones de reconocimiento, o hacia arriba y abajo para información y ayuda. Cada sección realiza una función determinada que es declarada mediante voz sintetizada por el dispositivo. Así se guía al usuario invidente para orientarlo en que sección se encuentra y que hace, las opciones, la ayuda completa e información adicional. Las secciones enlazan unas con otras en un ciclo de navegación.

La activación de las funciones se realiza mediante un toque continuo en la pantalla del dispositivo, que es reconocido emitiendo una vibración y con tonos cortos espaciados mientras está activada la función. Para contrarrestar las posibles condiciones de baja iluminación y sombras, siempre que sea necesario se activa automáticamente la iluminación del dispositivo.

Otras interacciones limpian la memoria, hacen decir a la aplicación que funcionalidad está lista o copia a memoria direcciones web de referencia.



Figura 7: Interacción durante el reconocimiento de billetes

3.2 Interacción mediante comandos de voz

Los comandos de voz se pueden usar para interactuar con el dispositivo y acceder a varias funciones, como la cámara, el calendario y la configuración. También se pueden utilizar para realizar llamadas y enviar mensajes sin necesidad de ver la pantalla del teléfono, lo que puede ser especialmente útil en situaciones de emergencia. Esto puede facilitar que las personas

con discapacidad visual usen su teléfono de forma independiente.

La aplicación posee esta capacidad de reconocer y ejecutar comandos de voz para cambiar entre funciones y activarlas o pedir información. Con esto permite a las personas discapacitadas visuales trabajar con la aplicación sin necesidad de interactuar manualmente. También provee varias opciones adicionales solo accesibles mediante los comandos de voz. Esta capacidad se mantiene activa aún esté la aplicación en un segundo plano o el dispositivo se encuentre con la pantalla apagada, lo que permite dar comandos sin necesidad de manipular el dispositivo.

Los comandos de voz están compuestos de una palabra de activación, y una o varias palabras de mando. Las palabras de activación de comando de voz son en este caso variantes de “*asistencia*”, por ejemplo, *asistencia*, *asiste*, *asistir*, *asista*. Si la aplicación reconoce que se ha dicho una palabra de activación, emite unos tonos indicando que esta lista para recibir el comando. Algunos ejemplos de palabras de mando establecidas son:

“*reconocer pesos*”, cambia a la opción de reconocimiento de billetes de moneda cubana.

“*activar pesos*”, similar al ejemplo anterior, cambia a esta opción, pero adicionalmente la activa.

“*hora*”, “*fecha*”, “*batería*”, la aplicación dice el estado de estos parámetros.

“*minimizar*”, lleva a segundo plano la aplicación.

La aplicación espera solo un corto tiempo por las palabras de comando. Si en este tiempo no se le ha dicho algo reconocible, emite unos tonos indicando que ha finalizado la espera, y el usuario necesitara decir de nuevo una palabra de activación.

3.3 Realimentación

Los usuarios con discapacidad visual necesitan una realimentación de las interacciones que realizan en el dispositivo y respuestas a las acciones de una manera consistente y fácilmente reconocible. Esta realimentación se realiza fundamentalmente de manera sonora, lo que incluye la síntesis de voz, de manera táctil, e incluso lumínica para las personas de baja visión.

Mientras están activadas en uso algunas operaciones, la aplicación emite tonos pausados de manera continua, así el usuario invidente se mantiene alerta sobre su actividad. Otros tonos anuncian el inicio y fin de las operaciones.

Un sistema Texto-a-Habla (TTS, *Text to Speech*) permite a las personas con discapacidad visual interactuar con dispositivos como computadoras o móviles. La versión 1.6 de Android introdujo características de asistencia, donde destaca la capacidad TTS (Google, 2009), con modelos y vocabularios

incorporados a los dispositivos por los fabricantes o descargables. También pueden incorporarse sintetizadores diferentes, según la preferencia de los usuarios del sistema. El uso de estos sintetizadores de voz permite que todas las guías y respuestas de la aplicación sea vocalizada por el motor TTS activo en el dispositivo.

La ayuda e información fundamental la aplicación la dice mediante el sistema de voz sintetizada (la información completa aparece en la manual de de usuario).

Si la aplicación entregara las respuestas tan pronto las obtiene, como puede ocurrir en dispositivos rápidos, estas se solaparían confusamente. Por esto solo se responde cuando ya se ha entregado una respuesta anterior y se aplica una pequeña pausa entre respuestas.

Algunas alertas, como los fallos en el reconocimiento o mala manipulación, son alertados por voz especificando el problema y/o recomendando al usuario como corregirlo.

Una recomendación cuando se emplea la aplicación como asistencia de navegación en exteriores, o en entornos con cierto nivel de ruido, es usar audífonos transductores, que transmiten el sonido a través de vibraciones, sin tapar los oídos ni atenuar los sonidos ambientales, útiles y necesarios necesarios para las personas ciegas para orientarse y comprender el entorno.

4. Pruebas, despliegue y recepción

Las pruebas de usuarios son una parte importante del proceso de desarrollo de aplicaciones para personas con discapacidades visuales. Implica trabajar con usuarios de este tipo para probar la interactividad, las funcionalidades, la reactividad a las respuestas, y obtener comentarios sobre cómo mejorar la accesibilidad y la usabilidad, además encontrar posibles errores. Estos comentarios se utilizan para realizar mejoras en el diseño de la interfaz de usuario, la interactividad y realizar correcciones de errores.

Las pruebas fueron realizadas por varios usuarios voluntarios, ciegos, de baja visión y videntes. Los dos primeros tipos de usuarios son avezados en el uso de los dispositivos móviles mediante sistemas de asistencia y poseían instalados y activados en sus dispositivos diferentes aplicaciones lectoras de pantalla, lo que permitió comprobar la compatibilidad de los mecanismos de interacción táctil y de síntesis de voz. Los problemas de incompatibilidades y conflictos se han ido resolviendo en lo fundamental en las versiones evolutivas de la aplicación.

Se comprobó además la compatibilidad con dispositivos, evaluándose en más de 20 diferentes, el más antiguo lanzado en 2014. Esto es importante pues asegura en la medida de lo posible encontrar problemas e incompatibilidades dentro de la gran variedad de dispositivos Android y sus fabricantes, que se deben corregir. También se cubrió un amplio rango de potencias de cómputo para evaluar la inmediatez de las

respuestas. Todo esto tiene en cuenta que en el sector poblacional de personas con discapacidades visuales tiene un sesgo hacia menores recursos económicos lo que implica poseer dispositivos más antiguos y/o de menores prestaciones.

Los usuarios reportaron los problemas iniciales, con detalles como que tipo de operación realizaban, que software adicional empleaban y las características de sus dispositivos. También realizaron recomendaciones de cambios y mejoras para mejor adaptación a sus necesidades y formas de uso de los dispositivos.

La aplicación es puesta a disposición de los usuarios en una web dedicada, a través de artículos explicativos y enlaces de descarga, y mediante un canal en el sistema de mensajería Telegram (OmenTec, 2023), al que se adjunta un chat para realimentación con los usuarios, donde estos plantean dudas, problemas y recomendaciones. También se distribuye mediante Apkls (Apkls, 2023), la plataforma nacional de distribución de aplicaciones para Android.

La aplicación previa (qBillete) pudo ser evaluada en su alcance y aceptación con ayuda de encuestas y recopilando opiniones de primera mano de los usuarios, con una puntuación de 4.48 sobre 5 puntos, lo que es un 89.6% de aceptación. Sin embargo, aSISTA aún está en despliegue limitado y en proceso de recopilación de datos, no se conoce el nivel de aceptación, aunque debe ser igual o mayor. Esto es algo a evaluar en profundidad para próximos trabajos. Quienes descargan e instalan la aplicación en sus dispositivos pueden a su vez intercambiarla por otras vías con otras personas.

Los comentarios y opiniones de los usuarios recibidas por diferentes vías muestran una gran aceptación de la aplicación, e intercambios con funcionarios de la ANCI permiten conocer que se ha logrado una amplia distribución de las aplicaciones previas entre los usuarios potenciales en todo el país, así que esta nueva no debe ser diferente. Al ayudar con el desenvolvimiento diario, cosa sensible y a veces complicada, los usuarios manifiestan agradecimiento al tenerse en cuenta sus necesidades y que les faciliten la inclusión social e independencia.

5. Conclusiones

La aplicación aSISTA se ha desarrollado para los discapacitados visuales como usuarios principales. Para ellos representa una ayuda a su independencia y desenvolvimiento

de la cotidianidad, así como una mejoría en su inserción y aporte social, ayudando en la inclusión de este segmento poblacional. Las opiniones han sido muy buenas y se ha mantenido una continua realimentación, con reportes de problemas, recomendaciones y soluciones, con los desarrolladores y entre los usuarios.

La aplicación muestra una buena efectividad en las operaciones de reconocimiento si se siguen las pautas de manipulación. La compatibilidad está mayormente asegurada con dispositivos lanzados en los últimos 7–8 años. La interactividad está hecha a la medida de las necesidades y posibilidades de las personas ciegas o de baja visión, facilitando su uso y así como la comprensión de las respuestas.

aSISTA es una herramienta en evolución y varios aspectos se han tenido en cuenta para su desarrollo futuro, mejorar sus capacidades actuales e incluir nuevas, tales como:

- Unificación del reconocimiento de billetes de las diferentes monedas y de las tarjetas mediante el empleo de una red neuronal (al momento de escribir este trabajo, es algo hecho y desplegado en aSISTA).
- Reconocer textos más amplios y en documentos.
- Mejor descripción de escena.
- Alerta ante obstáculos y posibles colisiones durante la marcha con ayuda del sensor de aceleración.
- Navegación y localización con ayuda de etiquetas (similar al primer punto en este momento) y GPS.
- Facilitar la localización y el reconocimiento de códigos QR, pues los sistemas en uso común se les hacen difíciles de usar a los ciegos.
- Reconocimiento y descripción de personas.
- Nuevos comandos de voz con más acciones.
- Extender a otras áreas de asistencia y discapacidades, como el reconocimiento de lenguajes de señas.

En general, aSISTA se convierte en una poderosa herramienta para las personas discapacitadas visuales al proveerles con la habilidad de identificar e interactuar con su entorno en maneras que antes les era muy difícil o imposible.

Referencias

- World Health Organization, (2012), Factsheet on visual impairment and blindness, world health organization, World Blind Union, <http://www.who.int/mediacentre/factsheets/fs282/en/>
- World Health Organization, (2019). World report on vision. Executive Summary. Department of Non communicable Diseases.
- Bourne, R. R. A. et al (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Global Health*, 5(9): e888–97

- Ecured, (2021). Asociación Nacional del Ciego. https://www.ecured.cu/Asociaci%C3%B3n_Nacional_de_Ciegos_y_D%C3%A9biles_Visuales
- Augustin, A. et al. (2007). Anxiety and depression prevalence rates in age-related macular degeneration. *Investigative Ophthalmology and Vision Science*, 48, 1498–1503
- Rodríguez, A., Nicolau, H., Montague, K., Guerreiro, J., Guerreiro, T. (2020). Open Challenges of Blind People Using Smartphones. *International Journal of Human-Computer Interaction*, 36:17, 1605-1622
- W3C, (2018). Web Content Accessibility Guidelines (WCAG) 2.1. W3C Recommendation 05 June 2018. <https://www.w3.org/TR/WCAG21/>
- W3C. (2023). Mobile Accessibility. <https://www.w3.org/WAI/standards-guidelines/mobile/>
- Sagale, A. A., Chaudhari, A. (2018). Assistive Technologies for Visually Impaired Users on Android Mobile Phones, *International Journal of Computer Applications*, Vol. 182 No. 32
- Dobosz, K. (2017). Designing Mobile Applications for Visually Impaired People. *Visually Impaired: Assistive Technologies, Challenges and Coping Strategies* (pp.103-126), Nova Science Publishers
- IDEAL (2023). Currency Identifier. <http://Ideal-group.org>
- LookTel (2023). Money Reader. <http://www.looktel.com/moneyreader>
- ViaOptaDaily (2020). <http://viaopta-apps.com/ViaOptaDaily.html>
- Mirwebsistem. Blind-Droid Wallet (2020). <http://droid.mirwebsistem.com/en/en-blind-droid-wallet/>
- EyeNote (2018). <https://www.eyenote.gov/>
- Google (2023a). Lookout. <https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.reveal>
- Microsoft Corp. (2023). Seeing AI. <https://www.microsoft.com/en-us/ai/seeing-ai>
- CloudSight Inc. (2023). TapTapSee. <https://desarrollarinclusion.cilsa.org/tap-tap-see/>
- Cruz, F., Yumang, A., Manalac, J., Canete, K., Milambiling, J. (2018). Smart backpack for the blind with light sensors, zigbee, rfid for grid-based election. *AIP Conference Proceedings*. vol. 2045, p. 020054. AIP Publishing
- Sato, D., Oh, U., Naito, K., Takagi, H., Kitani, K., Asakawa, C. (2017). Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. pp. 270–279. ACM.
- AlAbri, H. A., AlWesti, A. M., AlMaawali, M. A., AlShidhani, A. A. (2014). Naveye: Smart guide for blind students. *2014 Systems and Information Engineering Design Symposium (SIEDS)*. pp. 141–146. IEEE.
- Hossain, E., Rahman, M., Qaiduzzaman, K. (2020). Sightless Helper: An Interactive Mobile Application for Blind Assistance and Safe Navigation
- Be My Eyes (2023). <https://play.google.com/store/apps/details?id=com.bemyeyes.bemyeyes>
- BlindSquare, <https://www.blindsquare.com/about/>
- SuperSense. <https://play.google.com/store/apps/details?id=com.mediate.supersense>
- Salas, M. (2021). Asistencia a discapacitados visuales para reconocer billetes mediante aplicación móvil. *Revista Cubana de Ciencias Informáticas*, Vol. 15, No. Especial UCIENCIA II, septiembre, 2021, ISSN: 2227-1899, RNPS: 2301 (IX Taller Internacional de Software Libre. Android y sistemas operativos libres para móviles. Desarrollo de aplicaciones empleando tecnologías híbridas y/o nativas para móviles)
- Salas, M. (2022). qBillete. Asistencia en dispositivos móviles a personas con discapacidades visuales. *Memorias de la Convención y Feria Informática 2022. InfoTIC*. ISBN 978-959-7255-02-4
- Rashid, A., Prati, A., Cucchiara, R., (2013). On the design of embedded solutions to banknote recognition. *Optical Engineering* 52(9), 093106.
- Solyar, Z., Stubendek, A., Radvanyi, M., Karacs, K. (2011). Banknote recognition for visually impaired. *20th European Conference on Circuit Theory and Design*.
- Open Handset Alliance, (2023). Industry Leaders Announce Open Platform for Mobile Devices. http://www.openhandsetalliance.com/press_110507.html
- Statcounter MobileOS Global (2023a). <https://gs.statcounter.com/os-market-share/mobile/>
- Statcounter MobileOS Cuba (2023b). <https://gs.statcounter.com/os-market-share/mobile/cuba>
- Cubadebate, (2020). ¿Qué teléfonos móviles se usan más en Cuba?, <http://www.cubadebate.cu/noticias/2020/03/16/que-telefonos-moviles-se-usan-mas-en-cuba-infografia-y-top-10/>
- Google, (2023b). Android Studio, <https://developer.android.com/studio>
- OpenCV for Android, (2023). <https://opencv.org/android/>
- Gould, J.D., Lewis, C., (1985). Designing for usability key principles and what designers think. *Commun. ACM* 28(3), 300–311
- Qureshi, H.H., Wong, D.H.T., (2019). Requirements of a Mobile Application Design Model for Visually Impaired People. *Open International Journal of Informatics (OIJI)*, Vol. 7 Special Issue 2
- Qureshi, H.H., Hooi-Ten Wong, D., (2020). Usability of user-centric mobile application design from visually impaired people's perspective. *Universal access in human-computer interaction. Design approaches and supporting technologies. HCII 2020. Lecture notes in computer science*, vol. 12188. Springer 311–322

- Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M., (2016) . Computer vision for assistive technologies. *Computer Vision and Image Understanding* 154, 1–15
- NIST, (2023), The NIST Reference on Constants, units and Uncertainty, <https://physics.nist.gov/cuu/Units/units.html>
- dog-qiuqiu, (2021). Yolo-Fastest v1.1.0, Zenodo, <http://doi.org/10.5281/zenodo.5131532>
- Ooms, J., (2023). tesseract: Open Source OCR Engine. <https://docs.ropensci.org/tesseract/> (website) <https://github.com/ropensci/tesseract> (devel).
- Rogers, J., Sharp, Y., Preece, H., *Interaction design: beyond human computer interaction*, 6th Edition, Wiley, 2023.
- Google, (2009). An introduction to Text-To-Speech in Android. *Android Developers Blog*, <https://android-developers.googleblog.com/2009/09/introduction-to-text-to-speech-in.html>
- Nickolay, V. S., et al, (2020). Vosk Speech Recognition Toolkit: Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node. <https://github.com/alphacep/vosk-api>
- Suliman, N. A., Celsi, L. R., Li, W., Zomaya, A., Villari, M., (2022). Edge-Oriented Computing: A Survey on Research and Use Cases. *Energies*, 15(2), 452
- OmenTec Web (2023). <https://www.omentec.cubava.cu>
- Apklis (2023). <https://apklis.cu>