

Development of a Low-Cost Assistive Device for Scene Description for the Visually Impaired

Esther de Yeshua M. Pessanha, Victor A. Azevedo Costa Santos, Rigel P. Fernandes[✉], Thiago Silva de Souza

Abstract—This article proposes the development of a low-cost assistive device capable of capturing, analyzing, and providing auditory scene descriptions to help users understand and navigate their environment. The goal is to address the gap between the demand for assistive technologies and the availability of accessible devices in Brazil. The research follows an exploratory approach through the review of articles from Google Scholar and national statistical portals, considering publications from 2019 onward, along with the implementation of a functional prototype using accessible electronic components. The prototype demonstrates that it is possible to support the autonomy of blind people and people with visual impairments through a financially viable device. Experimental results show that the system achieves response times ranging from approximately 3 to 8.5 seconds, demonstrating its viability for real-time assistive applications.

Keywords—Accessibility; Artificial intelligence; Assistive technology; Low-cost devices; Raspberry Pi; Scene description; Visual impairment; Blindness.

I. INTRODUCTION

In Brazil, approximately 17.3 million people have some type of disability [1], with around 7 million affected by visual impairment [2], which impacts navigation, social interaction, safety, and other aspects of daily life that depend on a constant flow of visual information. The assistive technology sector emerges to mitigate these impacts, focusing on the development of products, equipment, devices, and services aimed at promoting functionality related to the activity and participation of people with disabilities or reduced mobility, seeking autonomy, independence, quality of life, and social inclusion [3].

The most traditional assistive technologies for visually impaired individuals include white canes, Braille, and optical magnifiers. However, with digital advancements, it has become possible to develop everything from screen-reader software to computer-vision devices integrated with Artificial Intelligence, offering features such as text reading, product identification, facial recognition, and scene understanding [4], [5]. The concept of Assistive Technology can be introduced through the following quote by Radabaugh: “For people without disabilities, technology makes things easier. For people with disabilities, technology makes things possible” [6].

Yet, access to assistive technologies remains a major challenge. In 2022, the World Health Organization (WHO), together with the United Nations Children’s Fund (UNICEF),

published a report revealing that more than 2.5 billion people need one or more assistive products [7]. However, almost 1 billion people do not have access to these technologies, especially in low- and middle-income countries, such as Brazil (an upper-middle-income country, according to the World Bank [8]), where access is extremely limited-reaching only 3% of those who need these products to live [7]. Previous research published by the WHO identifies lack of awareness, unaffordable prices, lack of services, inadequate quality, variety, and quantity of products, as well as challenges in procurement and supply-chain, as the main barriers [7]. These limitations are also widely discussed in the literature, which highlights issues related to usability, accessibility, and adoption of assistive technologies [9].

Given this scenario, this article proposes the development and documentation of a portable, low-cost, and high-performance assistive technology device designed to provide on-demand environmental descriptions for users with visual impairments, with the aim of increasing their autonomy.

The main contributions of this work are raising awareness of the challenges faced by individuals with visual impairments and encouraging the use of widely available technological components, as well as leveraging the growing accessibility of artificial intelligence resources to develop assistive technologies capable of generating meaningful social impact, while also serving as a cost-effective alternative to existing high-cost devices.

The remainder of this article is organized as follows. Section II presents the literature review and the existing market solutions for assistive technology devices, followed by Section III, which describes the proposed method and the development of the prototype. Section IV presents the experimental results, and finally, Section V presents the conclusion.

II. LITERATURE REVIEW

The demand for solutions aimed at increasing the inclusion of people with disabilities, especially in urban environments, is growing, strengthening the need to develop devices that support the independence of people with disabilities both inside and outside urban spaces. However, the development of such solutions is challenging not only due to the lack of investment in the sector but also because of the high cost of the final product, which, in most cases, makes access to these technologies unfeasible or difficult [10].

Assistive technologies for visually impaired individuals can be categorized according to their functionalities and working principles, including object detection, navigation, hybrid systems, and devices for activities of daily living (ADLs) [4],

Esther de Yeshua M. Pessanha, Victor A. Azevedo Costa Santos, Rigel P. Fernandes and Thiago Silva de Souza are affiliated with the Full Stack Development course at IBMEC-RJ (Centro Universitário IBMEC, Rio de Janeiro, Brazil). Emails: {eympessanha, victorvaacs, rigelfernandes, profthiagodesouza}@gmail.com.

[11]. The most common features found in existing devices include text-to-speech reading, object and person recognition and enumeration, and navigation [12].

It is possible to examine this scenario through similar wearable technologies already available on the market, designed to promote autonomy for blind and visually impaired individuals. In Brazil, this market is dominated by foreign solutions, mainly:

- OrCam MyEye (USD 2,761 / R\$ 14,900.00): A smart camera that magnetically attaches to eyeglasses. It can read text and recognize faces and products in real time, with offline processing [13]. In 2023, the Paraná state government purchased 147 units for students in the public school system, totaling USD 406,000 / R\$ 2.19 million [14].
- Envision Glasses (USD 1,899 / R\$ 10,252.70 - USD 3,499 / R\$ 18,891.10): A solution that integrates Envision AI software into a modern eyeglass frame. The device offers features such as scene description, text scanning, facial and product recognition, and even making phone calls. Its cost varies depending on the edition [15].

The analysis of similar technologies highlights the main barrier in the assistive technology sector: the cost of access. Cutting-edge solutions are imported products, priced in a way that makes them inaccessible to the vast majority of the Brazilian public.

III. PROPOSED METHOD

For the development of the proposed solution, a prototype-driven methodology was adopted, in which the design choices were guided by the use of resources readily available to the team, enabling the construction of a Minimum Viable Product (MVP) of a device aimed at increasing the autonomy of individuals with visual impairments. The main idea was to develop a device capable of capturing images, processing them using artificial intelligence, and providing an auditory description of the scene. Considering this concept and the tools initially available to the team, the MVP consists of components such as:

Component	Description
Raspberry Pi 4 (USD 75 / R\$ 385.46)	Used for image processing. A low-cost development board that runs Linux-based operating systems.
Air-conduction headset (USD 40 / R\$ 205.58)	Used for delivering audio output. It allows the user to listen to audio while maintaining awareness of surrounding environmental sounds since it transmits without obstructing the ear canal.
Smartphone	Used to capture images of the user's surroundings and transmit them to the Raspberry Pi for processing.
Google Gemini Pro API	Used to generate scene descriptions. Google's AI assistant, with the PRO version available free of charge for students (and also accessible at no cost through free-tier usage options).
Google Speech API	A text-to-speech service used to convert the scene description generated by the Gemini API into audio directly on the Raspberry Pi. The service can be accessed through Google's free-tier usage options.

TABLE I. Components used in the proposed MVP.

A. UML Diagram

The prototype's operational flow, illustrated in Fig. 1, begins when the user uploads an image through a web browser, using any device capable of accessing the interface. The image is transmitted via Wi-Fi to the Raspberry Pi, which uses an internet connection to forward the image to the Google Gemini API. The API analyzes the scene and returns a text description. This text is then converted into audio on the Raspberry Pi using a Google text-to-speech API, and the resulting audio is transmitted to the air-conduction headset via Bluetooth.

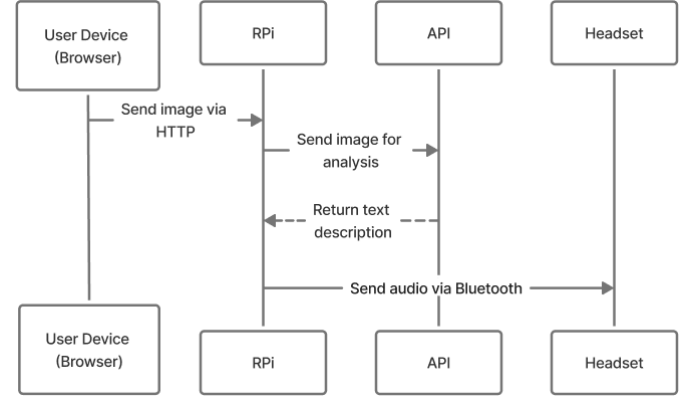


Fig. 1. UML Sequence Diagram

B. Implementation

The image processing script running on the Raspberry Pi was implemented in Python (Alg. 1). It starts a Flask server that enables users to upload an image via a web browser, processes the image, and returns an audio description. Additionally, a second script was implemented to automate the server startup process (Alg. 2), ensuring that the required services are initialized automatically when the device is powered on.

Algorithm 1 Image-processing script - app.py

```

1: Initialize Flask server
2: while server is running do
3:   Wait for HTTP request with captured image
4:   Receive image
5:   Send image to Google Gemini API with the prompt
6:   Receive textual description
7:   Convert text to speech using Google Speech API
8:   Play audio via Bluetooth headphones
9: end while
    
```

Algorithm 2 Startup automation script - boot.sh

```

1: Connect to the Bluetooth headset (device MAC address)
2: Activate the Python virtual environment (venv)
3: Execute python3 -u app.py
4: Initialize the Flask server
5: while the server is running do
6:   Wait for incoming HTTP requests
7: end while
    
```

IV. RESULTS

To evaluate the performance of the proposed system, a series of experiments were conducted using images with varying levels of visual complexity. The goal was to analyze the response time of the system when generating auditory descriptions using the Google Gemini API.

The test dataset was composed of images categorized into five levels of complexity: very low, low, medium, high, and very high. The complexity level of each image was determined according to the following criteria:

- Very low: one or two prominent objects, plain background, no textual information.
- Low: few objects with limited interaction and simple background.
- Medium: several objects, moderate background detail, and some contextual relationships.
- High: many objects, cluttered scenes, presence of text or signs, and multiple interactions between objects.
- Very high: crowded or highly detailed scenes containing numerous objects, significant contextual information, textual elements, and complex spatial relationships.

The results demonstrated in Fig. 2 indicate a clear correlation between scene complexity and response time. Simpler images required less processing time, while more complex scenes led to increased response latency.

Images with very low complexity were processed in the shortest time, as they contained fewer elements to be analyzed. In contrast, images classified as high and very high complexity required more processing time due to the increased number of objects and contextual relationships that needed to be interpreted. Although the measured response time also includes network communication and API processing overhead, the results indicate that image complexity remains an important factor influencing the overall system latency.

Despite this increase, the response time remained within acceptable limits for real-time assistive applications, indicating that the proposed system is suitable for practical use while maintaining consistent performance across different levels of scene complexity.



Fig. 2. Response time as a function of scene complexity.

V. CONCLUSION

This article seeks to encourage the development of low-cost assistive technologies as a means to increase the autonomy and inclusion of people with visual impairments. The prototype's experimental results show that the system provides scene descriptions within a response time suitable for assistive use, ranging from approximately 3 seconds for simple images to 8.5 seconds for more complex scenarios. These results demonstrate that the proposed solution is capable of handling different levels of visual complexity while maintaining acceptable performance, while requiring an estimated development cost of only USD 115 / R\$ 590.00.

Planned improvements include replacing the browser-based image capture system with an ESP32 equipped with an integrated camera module, migrating to local processing to eliminate dependence on network connectivity, replacing the air-conduction headset with a bone-conduction headset to keep the ear canal unobstructed, and enabling battery-powered operation to eliminate dependence on external power sources. Together, these improvements enhance portability, usability, situational awareness, and autonomy, strengthening the viability of low-cost assistive solutions within the Brazilian context.

REFERENCES

- [1] Agência IBGE Notícias, "National health survey 2019: The country has 17.3 million people with some type of disability," 2021. [Online]. Available: <https://encurtador.com.br/Oxry>
- [2] IBGE, "Pesquisa nacional de saúde – tabela 8204: Pessoas com deficiência visual," 2019. [Online]. Available: <https://sidra.ibge.gov.br/tabela/8204>
- [3] Governo do Brasil, "What is assistive technology?" 2022. [Online]. Available: <https://www.gov.br/saude/pt-br/assuntos/saude-da-pessoa-com-deficiencia/faq/o-que-e-tecnologia-assistiva>
- [4] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, 2018.
- [5] P. Kathiria, S. H. Mankad, J. Patel, M. Kapadia, and N. Lakdawala, "Assistive systems for visually impaired people: A survey on current requirements and advancements," *Neurocomputing*, 2024.
- [6] M. P. Radabaugh, "Study on the financing of assistive technology devices and services for individuals with disabilities," 1993.
- [7] United Nations Brazil, "Approximately one billion people with disabilities are denied access to assistive technology," 2022. [Online]. Available: <https://encurtador.com.br/LPHb>
- [8] S. Mahdi, "Brazil and the middle-income trap," 2024. [Online]. Available: <https://www.worldbank.org/pt/news/opinion/2024/09/12/brasil-e-a-armadilha-da-renda-media>
- [9] Z. J. Muhsin, R. Qahwaji, F. Ghanchi, and M. Al-Tae, "Review of substitutive assistive tools and technologies for people with visual impairments," *Journal on Multimodal User Interfaces*, 2023.
- [10] J. R. Beingolea, M. A. Zea-Vargas, R. Huallpa, X. Vilca, R. Bolivar, and J. Rendulich, "Assistive devices: Technology development for the visually impaired," *Designs*, 2021.
- [11] S. Zafar, M. Asif, M. B. Ahmad, T. M. Ghazal, T. Faiz, M. Ahmad, and M. A. Khan, "Assistive devices analysis for visually impaired persons: A review on taxonomy," *IEEE Access*, 2022.
- [12] G. de Queiroz Pereira, A. C. de Francisco, D. P. B. Renaux, and L. A. Pilatti, "Assistive technology tools for the acquisition of knowledge for the visually impaired," *Revista Tecnologia e Sociedade*, 2023.
- [13] P. Patriarca, "Óculos falantes: como um dispositivo de inteligência artificial ajuda estudantes cegos no brasil," 2022. [Online]. Available: <https://encurtador.com.br/CGVk>
- [14] Governo do Paraná, "Governo compra óculos de inteligência artificial para alunos cegos da rede estadual," 2023. [Online]. Available: <https://encurtador.com.br/YIKh>
- [15] Envision, "Envision glasses," [Online]. Available: <https://www.letsenvision.com/glasses/home>