DESIGN OF A SMARTPHONE APPLICATION TO PROVIDE SHOPPING ASSISTANCE TO PEOPLE WHO ARE BLIND OR HAVE LOW VISION

BY

GREGORY LEMASURIER
B.S. UNIVERSITY OF MASSACHUSETTS LOWELL (2020)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
COMPUTER SCIENCE
UNIVERSITY OF MASSACHUSETTS LOWELL

MAY 2022
DESIGN OF A SMARTPHONE APPLICATION TO PROVIDE SHOPPING ASSISTANCE TO PEOPLE WHO ARE BLIND OR HAVE LOW VISION

BY

GREGORY LEMASURIER
B.S. UNIVERSITY OF MASSACHUSETTS LOWELL (2020)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE COMPUTER SCIENCE UNIVERSITY OF MASSACHUSETTS LOWELL

MAY 2022

Signature of Author: _______________________________ Date: 04/05/2022

Signature of Thesis Supervisor: _______________________________ Dr. Holly A. Yanco

Signature of Other Thesis Committee Members:

Committee Member: _______________________________ Dr. Mohammad Arif Ul Alam

Committee Member: _______________________________ Dr. Aaron Steinfeld
DESIGN OF A SMARTPHONE APPLICATION TO PROVIDE SHOPPING ASSISTANCE TO PEOPLE WHO ARE BLIND OR HAVE LOW VISION

BY

GREGORY LEMASURIER

ABSTRACT OF A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

UNIVERSITY OF MASSACHUSETTS LOWELL

MAY 2022

Thesis Supervisor: Dr. Holly Yanco
Professor, Computer Science
Abstract

According to the World Health Organization in 2021, approximately 2.2 billion people are blind or have visual impairments. People who are blind or have low vision (B/LV) rely on assistance from others as well as from assistive technology to complete many everyday tasks, which require visual sensory information.

We initially conducted a survey of devices for people who are B/LV which focused on the information communicated to the user through different feedback methods and the tasks these devices are designed to assist people with. After discovering a need for object finding devices in our survey, we distributed a questionnaire to local agencies for people who are B/LV. Our first questionnaire focused on learning about assistive devices that are commonly used by this population, as well as challenges that they face while shopping. From this questionnaire, we found that there is a need for a device to assist with various tasks while shopping, including navigating to desired products and reading information on labels. We then expanded upon the initial survey of devices to investigate the sensor packages used for indoor and outdoor sensing. Next, in a follow up questionnaire, we asked questions to gather feedback on our initial design plans for this device including how beneficial they thought it would be, how the device should communicate information to them, and any privacy concerns they had with sharing data to build a map shared between users.

The system that we designed and implemented uses a smartphone and its built in sensors to provide shopping assistance, primarily using speech. Labels are read to identify products and to provide answers to questions users may ask regarding the products. Simultaneous Localization and Mapping (SLAM) is used to map the stores and provide navigation instructions to users. These maps are community updated, where all devices contribute data to ensure that the maps are kept up to date. This project is open source\(^1\).

\(^1\)https://github.com/GregoryLeMasurier/ShoppingAssistanceForBlindAndLowVision
Acknowledgements

I would like to thank my advisor, Dr. Holly Yanco, for her guidance, patience, and support throughout the five years I have been in the Human Robot Interaction Lab. I have certainly learned a lot from her advice and guidance while working on this work and various projects in the lab. I would also like to thank my committee members Dr. Mohammad Arif Ul Alam and Dr. Aaron Steinfeld for being on my committee and for their suggestions and feedback on this work.

I would like to thank my labmates at the Human Robot Interaction Lab and NERVE Center, especially Jordan Allspaw and Zhao Han who I have had the pleasure to work with and have learned a lot from while working in the lab. Additionally, I would like to thank Brian Flynn and Adam Norton who helped set up the test arena used in this work. Thank you to all of my labmates who piloted and offered suggestions to improve the device. Thank you to everyone who proofread my paper or listened to my practice talks.

Finally, thank you to the Massachusetts Commission for the Blind, the Massachusetts Association for the Blind and Visually Impaired, and the Lowell Association for the Blind who helped recruit participants for our initial questionnaires.
# Table of Contents

Abstract  

Acknowledgements  

Table of Contents  

List of Tables  

List of Illustrations  

1 Introduction  

1.1 Problem Statement  

1.2 Thesis Contributions  

2 Related Work  

2.1 Printed Information Access  

2.2 Object Recognition  

2.3 Navigation  

2.3.1 Electronic Travel Aids (ETA)  

2.3.2 Electronic Orientation Aids (EOA)  

2.3.3 Position Locator Devices (PLD)  

2.4 Shopping Assistance  


# System Design Process

3.1 Literature Review .................................................. 13
3.2 Formative Questionnaire ........................................... 14
3.3 Sensor Packages ..................................................... 18
3.4 Proposed Device Feedback Questionnaire ......................... 20
3.5 Summary .............................................................. 22

# System Implementation

4.1 System Overview ..................................................... 24
4.2 Smartphone Selection ............................................... 25
4.3 User Input Commands ............................................... 26
4.4 ARCore .............................................................. 29
4.5 Database of Community Updated Maps ........................... 30
4.6 Localization .......................................................... 31
4.7 Label Parser .......................................................... 32
   4.7.1 Product Name ................................................... 33
   4.7.2 Nutrition Information .......................................... 34
   4.7.3 Allergy Information ............................................ 34
4.8 Product Identification .............................................. 36
   4.8.1 Dataset .......................................................... 36
   4.8.2 Object Detection Model ....................................... 36
   4.8.3 Non-Maximum Suppression .................................... 38
   4.8.4 Product Identification ......................................... 38
   4.8.5 Mapping Products .............................................. 38
4.9 Navigation ............................................................ 40
   4.9.1 Walkable Paths ................................................ 41
List of Tables

3.1 Design Considerations from Formative Survey ........................................ 18
3.2 Sensor packages used by 16 assistive devices that were evaluated in both indoor and outdoor environments ......................................................... 19
4.1 User Input Commands ................................................................. 28
4.2 Summary of Text Reading Assistance ........................................... 43
5.1 Detected cereal product names across different resolutions. ................. 60
5.2 Detected soup product names across different resolutions. ................. 61
List of Illustrations

3.1 Distribution of reported degree of visual impairment across participants who participated in our formative questionnaire. 15

3.2 Distribution of reported degree of visual impairment across participants who participated in our proposed device feedback questionnaire. 21

4.1 A flowchart diagram showing how the app operates. The product identifier and map update modules loop until the application ends. The user can interrupt these modules by pressing down on the screen, and then they enter assistive mode. In assistive modes that use the label reader, the product identifier and map updates are disabled. In assistive modes that find products, the A* path planning module generates a path to the desired product and then the application guides the user through the path while also starting the product identification and map update loop. 25

4.2 The Samsung Galaxy S21 phone used in this work. Camera A is the ultra wide camera, Camera B is the wide camera, and Camera C is the telephoto camera. 27

4.3 An example database entry including the origin anchor id, waypoints for walkable paths, and products. 31

4.4 Examples of product labels. On the left the larger text on the label consists of the brand name, followed by the zero sugar descriptor, followed by the product name. On the right the product label consists of a large size descriptor followed by a brand name, followed by a product name. 34
4.5 Examples of product nutrition facts labels. ........................................ 35
4.6 Examples of allergy information found on product labels. .................. 35
4.7 Example product detection results from our product detection model. Note, the bounding box was drawn four times thinner in this case compared to all other figures. We did so to make the overlapping detections more noticeable. .......... 37
4.8 Example product detection results from our product detection model after applying our NMS filter. ................................................................. 39
4.9 Example database entry for a detected product. ................................. 40
4.10 Anchor generated at the detected product location. The products on the left were detected while the user was walking. The products on the right were detected through the doors in a freezer aisle. ................................. 41
5.1 Examples of product detections, shown through the red boxes, on canned products, boxed products, and bottled products. ............................... 46
5.2 Examples of product detections, shown through the red boxes on the bottom center of the shelf, on produce and thin horizontally laid bagged food. ............... 47
5.3 Examples of incorrect product detections, shown through the red boxes, on reflections and products with labels that include other products on their logo, partially occluded products, and products that are grouped together by something such as plastic six pack rings. ......................................................... 48
5.4 Example of product detections, shown through the red boxes, on a close view of a cluster of products. Note that most products are not identified at this closer range. 49
5.5 Examples the shelf view (left), cluster of products view (middle), and single product view (right). ................................................................. 50
5.6 Example of canned products on the shelf, where the products’ logos are not all facing forwards. ......................................................... 50
5.7 Example of a product with text that is written in a circle. . . . . . . . . . . . . . . 51
5.8 Example of products that are partially occluded by shelf tags and other items being
sold on a hang tag. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
5.9 Example of products with stylized logos. . . . . . . . . . . . . . . . . . . . . . . . 53
5.10 Example of a label where nutrition fact key and value are split across two lines. . . 54
5.11 Example of a label where the calories are labeled with English and Spanish keys. . 54
5.12 Most common example of a label where nutrition facts are split across multiple
columns. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
5.13 Variation of a label where nutrition facts are split across multiple columns. . . . 56
5.14 Example of the resulting text (right) read aloud to the user when using the “Read”
command on the scene (left). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
5.15 The experimental set up for our planned user study. . . . . . . . . . . . . . . . . . 58
Chapter 1

Introduction

1.1 Problem Statement

People who are blind or have low vision, which we will abbreviate as B/LV in this manuscript, have to rely on assistance from others as well as from assistive technology to complete many everyday tasks which require visual sensory information. According to the World Health Organization, approximately 2.2 billion people are blind or have visual impairments as of October 2021 [1]. This number includes mild visual impairments that are can be corrected with prescription glasses. In 2015, a study conducted by Bourne et al. found that approximately 36 million people were completely blind and 216 million people had moderate to severe visual impairments [2]. Despite investments to prevent visual impairments, the age structure and growth of the world population has caused an accelerating increase in the number of cases of visual impairments in the world [2]. Thus, it is necessary to provide assistive solutions that enable this ever growing population to be capable of independently and efficiently completing every day tasks.

Assistive solutions for people who are B/LV can be categorized into one of three subcategories: vision enhancement, vision replacement, and vision substitution, as described by Dakopoulos and Bourbakis [3]. Vision replacement requires surgery, which is invasive and not always an af-
fordable option. Vision enhancement includes devices which express information in visual means for people with low vision; therefore, these devices would not work for people who are blind. Vision substitution devices have been proven to be effective by Haigh et al. [4], who showed that naive blindfolded users could use this type of device to accurately identify the direction that a letter “E” was facing. These devices are noninvasive and are relatively cost friendly compared to vision replacement methods. Thus, visual substitution devices should be designed to assist people who are B/LV in completing everyday tasks that require visual sensory information, such as shopping.

Shopping requires people to use visual sensory information to identify and find products in the store, access important information on labels and store signs, safely navigate around the store, and to enter in their pin number when checking out. This is especially difficult for people who are blind or have severe low vision.

While some stores have braille signs located around the store to help people who are B/LV traverse the store, such as on the ends of aisles or on doorways, not all stores provide this assistance [5]. Additionally, many individuals who are B/LV shop with others, such as family, friends, or get help from store employees [5, 6]. Smartphone applications have also been developed to connect people who are B/LV with a sighted volunteer. These volunteers can help assist a user by looking at images from the user’s camera. Commercial applications that provide this service include Be My Eyes¹ and Aira².

While these options provide sufficient shopping assistance, they require the individual to be dependent on another person. Lee et al. found that while people who are B/LV often shop with others, they ultimately prefer not to, mostly because they feel uncomfortable asking others for assistance [6].

Assistive technology is essential to enable independent shopping for people who are B/LV. There are many different options available to provide different types of assistance while shopping.

¹https://www.bemyeyes.com/
²https://aira.io/
For navigation around stores, two popular solutions involve using a white cane and or a guide dog, however guide dogs are expensive [5]. Though these solutions are convenient, they provide people who are B/LV with very limited information about the world around them. Both options also require that the user has previously visited and has learned the layout of the store before.

Commercial products also exist that provide assistance to identify objects or products. For example, Seeing AI\textsuperscript{3} reads barcodes to help identify products. Another smartphone application, Google Lookout\textsuperscript{4}, uses machine learning techniques to read text and identify objects in the user’s environment.

While all of these assistive technology options enable a user to independently complete a particular task, such as avoiding obstacles while walking, navigating to areas in a store, or identify objects, it is especially important for an assistive device to be capable of providing all necessary assistance when users are shopping at a store they are unfamiliar with. When B/LV individuals are shopping at a new store, they would not have an understanding of the layout of the store. Uribe-Fernández et al. found that B/LV individuals are more comfortable shopping on their own after they have learned the layout of the stores [5].

To enable people who are B/LV to be comfortable while independently shopping at stores they are unfamiliar with, there is a need for an assistive device that can provide assistance walking to desired aisles, retrieving information on product labels, and identifying products.

1.2 Thesis Contributions

In this work, we present the design of a prototype assistive device to provide shopping assistance to people who are B/LV. To design this assistive device, we first distributed a formative questionnaire to identify what assistance would be beneficial to people who are B/LV while shopping. Then we build off of our initial literature review to investigate the sensor packages used by assistive devices.

\textsuperscript{3}https://www.microsoft.com/en-us/ai/seeing-ai
that were tested in both indoor and outdoor environments. Next we created initial design plans for an assistive device and distributed a second questionnaire to gather feedback on the initial device design. The final device design was determined based on the findings of the literature review and the two formative questionnaires, keeping our target user group involved throughout the process of designing the device.

Then we developed an Android based smartphone application to enable people who are B/LV to shop in stores, including stores that they are not familiar with. This application enables a user to shop in stores by providing assistance with navigation to desired products, reading information on product labels, and identifying products.

To answer questions users have regarding products, we developed a label parser which was built off of Google’s Optical Character Recognition (OCR) - ML Kit Vision API. This parser is able to extract various information that is necessary to provide shopping assistance such as the product name, nutrition information, and allergy information.

Additionally we have developed a robust product identifier. This product identifier is developed using two different machine learning modules. First, products are recognized using a YOLO-v5 [7] model that was trained on the SKU-110K dataset [8]. Then we identify the product name using the label parser we have developed. This approach is robust and does not require training data for each specific product that we would like to identify, unlike traditional single model object identification approaches.

ARCore is utilized to track the user’s location in a store based on the device’s internal IMU sensor, as well as from features in the environment. Using ARCore, we generate labeled maps that include walkable paths. With these maps, we are able to generate routes for a user to follow to their desired product using an A* algorithm for path planning.

The maps generated with this application are community updated using a Firebase database.

---

5https://developers.google.com/ml-kit/vision
6https://developers.google.com/ar
This allows us to keep up to date maps, as product and walkable locations are updated online as all users use the app.

Users can get assistance while shopping by using a set of spoken command options. The application then uses a combination of speech and audio tones to communicate information back to the user.

We have conducted a proof of concept evaluation for the prototype device’s functionality. While the device is still in the prototype phase, the results show that this approach is valid and can be further improved to create a fully functioning assistive application. Additionally, we describe our plans for a user study, which we were unfortunately unable to recruit participants for at this time, likely due to the pandemic.
Chapter 2

Related Work

Assistive technology for people who are blind or have low vision (B/LV) has been categorized into five branches of research: mobility, navigation, object recognition, printed information access, and social interaction by Bhowmick and Hazarika [9]. Bhowmick and Hazarika’s claim is supported by the findings of Brady et al. who identified and categorized areas in the daily lives of people who are B/LV where participants asked for assistance through an app [10]. This work focuses on providing shopping assistance, which falls into the categories of printed information access, object recognition, and navigation.

2.1 Printed Information Access

There are many different methods to provide printed information access to people who are B/LV. Providing access to printed information is essential when providing shopping assistance in order to communicate information found on product labels or signs in the store. Finger-mounted text readers [11, 12] have been developed to assist people who are B/LV access text based information found in newspapers, books, or menus. These text readers use a camera mounted on the user’s finger to read the text to which that the user is pointing. Several different feedback methods were explored in these text readers: vibrotactile [11], speech [11], and sonification [12] feedback was
used to indicate the direction that the user should move their finger. Both devices read the text below the user’s finger aloud.

Some work has also investigated converting graphical information from computer screens, such as graphs, to sonified audio patterns [13]. Using this device, users were able to learn the tone patterns outputted by this device to describe graphics.

Lee et al. designed a set of description guidelines for describing products to people who are B/LV [6]. Their proposed device is a wearable mixed reality headset that has a camera built in. The product is then recognized and looked up in a product database. The device then conveys a product description through speech.

Research groups have also designed assistive devices to assist people who are B/LV with reading product labels. Yi et al. designed a wearable assistive device that included a camera mounted to glasses [14]. To identify the region of interest and to segment the product out of the scene, users would shake the product. Text is then extracted from the product label using an off the shelf OCR system. The device then uses speech to communicate the text on the label to the user. Several other label reading assistive devices for people who are B/LV also have a similar design [15]. These devices all read text from labels by passing an image from a camera to an OCR system which then conveys text found on the product to the user through speech.

Extracting information from product labels is not limited to the domain of assistive technology for people who are B/LV. Nutrition label parsing has also been used to improve people’s diets [16, 17]. Grubert and Gao investigated passing images from a smartphone through OCR and then recognizing words using Tesseract, template matching, and multi-class SVM classification algorithms [17]. Similarly, Matsunaga and Sullivan extracted nutrition name and value pairs using OCR and Tesseract [18]. Another approach to extracting information from nutrition labels involves using OCR and deep learning[16]. Our work follows a similar process as these papers, however, we use Google’s Optical Character Recognition (OCR) - ML Kit Vision API, which has direct Android support. Additionally, we define a set of simple rules to parse these standardized
product labels rather than applying machine learning techniques. This helps reduce the number of computations required to parse a label, potentially leading to faster in app performance.

2.2 Object Recognition

Object recognition is essential for an assistive device that provides shopping assistance so that the device can identify and distinguish products.

One set of state of the art object detector comes from the YOLO family of object detection models [19]. These models are fast and are used to detect objects in image data. Our product detector uses the latest detector in the YOLO family, YOLOv5 [7]. A device has been designed that uses the deep learning to analyze distortions of projected green light to identify objects [20]. Methods such as the one proposed by Lowe also exist to detect products from 3D feature data [21].

Several devices have been designed that provide object recognition assistance for people who are B/LV. VizWiz is a smartphone app where users can submit a photo and corresponding question that will be answered through crowdsourcing [22]. Users asked object recognition related questions such as “What can is this?”, “What kind of drink is this?”, or “Which can is the corn?”. Similarly, VizLens enables users to submit an image which is then labeled through a crowdsourcing program [23]. Users can then touch an area on the image to hear the area’s corresponding labels. VizWiz::LocateIt uses both crowdsourcing and computer vision to identify objects [24]. Images that were crowdsourced were labeled. Audio tones with modifications to frequency and pitch and speech were used to communicate the distance to the user’s desired object. While crowdsourcing might result in higher accuracy, since people are identifying objects and answering questions, and requires no training data, this method still requires a person who is B/LV to be dependant on someone else. Our object identification module uses machine learning and text parsing to enable people who are B/LV to shop independently.

Groups have also investigated methods for product recognition in stores. Tonioni and Di Ste-
fanomen developed a system to recognize products and compare their locations to an expected layout [25]. Feature mapping and one-shot learning using deep learning methods has also been investigated to identify products [26].

2.3 Navigation

Two popular solutions for assistive navigation involve using a white cane and or a guide dog [5]. These devices are popular and convenient, however, they provide limited information about the world around them to their user and they require a user to have already visited and learned the layout of the area. Assistive devices that provide navigation assistance can be classified as either Electronic Travel Aids (ETAs), Electronic Orientation Aids (EOAs), or Position Locator Devices (PLDs) [27].

2.3.1 Electronic Travel Aids (ETA)

Dakopoulos and Bourbakis defines Electronic Travel Aids as “devices that transform information about the environment that would normally be relayed through vision into a form that can be conveyed through another sensory modality” [27]. This includes information such as surrounding objects or locations of obstacles.

Many different forms of ETAs have been developed throughout the history of the field of assistive technology for people who are B/LV. Wearable belts have been developed to inform the user of surrounding obstacles through vibrotactile feedback [28]. ETAs can also take the form of a cane. For example, the Intelligent Cane is a portable ETA that uses vibrotactile and sonification feedback to communicate the distance of obstacles to its user [29]. Additionally, an assistive suitcase has been designed to communicate when it predicts that a person is likely to collide with the blind user [30].
2.3.2 Electronic Orientation Aids (EOA)

Electronic Orientation Aids are defined by Dakopoulos and Bourbakis as “devices that provide orientation prior to, or during the travel” [27]. These devices signal to a user how they should navigate through following a path or providing other information about the user’s orientation [31].

Several groups have investigated using robots as EOAs to provide navigation assistance to people who are B/LV [32, 33]. Other groups have designed smartphone based EOAs to provide navigation instructions [34, 35, 36].

Several vibrotactile belts have also been developed. Nagel et al. designed a belt which indicates the direction of the north pole; this allows the user to understand their relative orientation [37]. Additionally vibrotactile belts have been used to communicate the direction that a user should travel [38].

2.3.3 Position Locator Devices (PLD)

Position Locator Devices are defined by Dakopoulos and Bourbakis as “technologies like GPS, European Geostationary Navigation Overlay Service (EGNOS), etc” [27]. These devices provide users with information about where they are currently located, such as a street name.

An example of a PLD is ASSIST [35], which not only provides navigation instructions but also informs the user when they have approached an area of interest in indoor settings. Another PLD uses electro-tactile gloves to communicate the surrounding environment and important landmarks to the user [39].

2.4 Shopping Assistance

Several different methods to provide shopping assistance for people who are B/LV have been explored. Kulyukin and Kutiyanawala developed a set of design guidelines for assistive shopping systems [40]. Robots have been used to guide people who are B/LV through stores [33]. Some
groups have also investigated using RFID for navigation [41, 33, 42] and to identify products [43]. QR codes have also been used to identify products [41]. These options require modifications to the store environment to function. In addition to providing navigation assistance, groups have also investigated providing directional guidance to assist the user in reaching for their desired product on a shelf [44].

Google Lookout is a smartphone application that can scan product labels and attempts to find a match in a product database. This app can also identify products by scanning barcodes. Google Lookout informs the user about the brand, flavor and product name.

In addition to Google Lookout, several other devices have been designed that read barcodes to identify products [45, 24, 46]. ShopTalk is an assistive device that was designed to assist with both product identification and store navigation [46]. This device uses speech feedback methods to communicate navigation instructions as well as the location of the desired product relative to the user [47]. ShopTalk requires a map to be manually generated for each store. The map includes locations of aisles and product locations. Barcodes found on the shelves are used to identify products. While this would be a relatively low cost assistive device, users are unable to efficiently browse shelves as they would have to scan each barcode to do so. Additionally, the information this device can convey to the user is dependent on their product database.

With our work, we hope to contribute to the field by addressing these limitations through our community updated maps which are updated as the user walks around and our app scans the shelves. Additionally, we have created a product label parser which provides the user with information about the product when they request it.
Chapter 3

System Design Process

When designing an assistive device for people who are B/LV, there are many different design decisions to consider. First, we need to identify what problem we are trying to solve. Are we trying to provide assistance to navigate to a desired location? To help avoid obstacles? To read information conveyed through text? Each task would require different design decisions to be made to provide appropriate assistance. A device that helps a user read text will have a completely different design than one that specializes in helping a user safely travel by avoiding obstacles.

It is very important to involve the target population in the design of the device as early as possible. By involving the target population in early design decisions we ensure that the device we designed would fit the needs of the target population. For example, should the device be wearable (where the device is worn by the user), portable (where the device is held in the hand of the user), or should the device be located in the environment (such as a robot kiosk station [48])?

Assistive devices must use sensor packages which are dependant on the problem the device is attempting to solve and the desired interface type. When designing a device for people who are B/LV, it is also very important to consider the methods that the device should accept as user input and the feedback methods used by the device as output. These should be simple and intuitive for the target population. Many devices use sonification, speech, vibrotactile, or other methods.
of tactile feedback to convey information to the user [49]; it is necessary to involve the target population in the decision for feedback methods.

For these device considerations, we first conducted a literature review analyzing the tasks devices were used for and what feedback methods were used to convey information back to the user [49]. Then we conducted a formative questionnaire to identify if there was a need to provide shopping assistance and to gather initial preferences on design decisions for an assistive device that provides shopping assistance. Then we expanded our literature review to investigate the sensor packages used in assistive devices. Next, we created an initial design for a device to provide shopping assistance. Finally, we distributed a follow up questionnaire to gather initial feedback before beginning to develop a prototype of the device.

3.1 Literature Review

As a start, I wrote a survey paper which analyzed 97 assistive devices for people who are B/LV [49]. This work was expanded by adding four additional devices that we learned of after the initial survey. These additional devices bring the total of analyzed devices to 101 assistive devices for the B/LV. Through this survey, we found that most assistive devices for the blind were designed to assist with navigation tasks. Of the surveyed devices, 87 focused on navigation, 9 on information gathering, and 5 on object finding. Thus, we saw an opportunity to help further develop the field by focusing on providing assistance in information gathering and object finding. A scenario that encompasses all three of these groups of devices would be a shopping task, where people would need to navigate around the store, get information from labels and signs, and find the objects that they are looking for on the shelves.

Furthermore, through this survey we found that the information conveyed through different feedback methods is not consistent across devices. Thus, it is important to identify what feedback methods people prefer, as well as what information they would like to be conveyed. This shows
the importance of including end user involvement to keep up with best practices when designing a device for people with disabilities.

### 3.2 Formative Questionnaire

It is essential to involve the target population in the design of an assistive device as early as possible. To identify if there is a need for shopping assistance assistive devices for people who are B/LV, and to identify important aspects of the device design, we designed and distributed an online survey to people who are B/LV.

This formative questionnaire was conducted online through Qualtrics\(^1\) and was approved by the university’s IRB. This questionnaire was conducted online due to the pandemic, which likely resulted in a sample bias, as the participants who responded to participate in our questionnaire are more likely to use technology. The questionnaire was expected to take less than one hour to complete. As a compensation for the participants’ time and effort we sent out $15 Amazon gift cards to everyone who completed the questionnaire.

To recruit participants, we reached out to local organizations including the Massachusetts Commission for the Blind, the Massachusetts Association for the Blind and Visually Impaired, and the Lowell Association for the Blind.

A total of 16 participants completed our formative questionnaire, which can be found in Appendix A. The participants’ ages ranged from 18 to 67 ($M = 32.46, SD = 15.70$). Ten participants identified as male, five as female, and one non-binary. Eight participants had their visual impairment since birth, one since age 1, one since age 2, two since age 3, one since age 8, one since age 17, one since age 33, and one since age 64. A distribution of our participants’ degrees of visual impairment can be seen in Figure 3.1.

To gather information regarding the shopping habits of our participants, we first asked questions regarding the participant’s shopping frequency, what types of stores they shop at, and who

\(^1\)https://www.qualtrics.com/
they shop with. We found that three participants would shop at stores once every couple of weeks, six shopped once a week, five more than once a week, one less frequently than a month, and one selected that they never shop. Participants reported shopping in stores including grocery stores (10), general retail stores (9), and convenience stores (4). We then asked if participants if they shop with others; three participants indicated that they would shop with friends, two selected that they do not shop, six shop alone, and five with family.

Next, to learn about areas that people who are B/LV would benefit from a device to assist them with shopping, we asked participants about challenges that they faced while shopping at a new store. We focused on challenges faced in new stores, as in this scenario they would be unfamiliar with the store layout. The participants reported that the major challenges while shopping at a new store include finding desired products (9) and not knowing the floor plan or layout of the store (9). Additionally, participants reported trouble with labels and text on products or store signs (3). One participant reported that they have “anxiety because [they] start to stress about people staring or wanting to see the items [they are] in the way of.” Finally some participants also

Figure 3.1: Distribution of reported degree of visual impairment across participants who participated in our formative questionnaire.
reported challenges with the checkout pinpads.

We also asked participants to rate on a 5 point Likert-type scale, how easy they consider finding desired items at a new store. Seven participants reported that finding items in a new store is indifferent, eight selected difficult, and one chose very difficult. These results show the need to provide assistance to help users navigate around stores to find their desired products and to read text found on products or store signs.

We next asked questions to learn about assistive devices that participants were aware of or have used to assist them with shopping. Six participants had used assistive devices before. Some examples of these devices include iPhone apps, barcode scanners, cameras, magnifiers, Seeing AI, Aria, and color readers. Out of these six participants, four agreed that their device is easy to use, and two strongly agreed. Participants reported that they use their device to read text based information, such as information found on product labels.

The participants reported that some limitations of these devices include performance in varying lighting conditions, difficulties focusing the camera, difficulties finding a label or specific text, inaccuracies in the device, and magnifiers that are not strong enough to be clear to the user. Additionally one participant reported that a limitation of their device was that it was unable to guide them to the locations of items, registers, or exits.

We then asked participants about what feedback methods their device uses. Three participants reported that their assistive devices provide feedback through voice or speech, one through text, and two through magnification. Providing feedback through text and magnification might be sufficient for some people who have low vision; however, it is necessary to use other means of feedback modalities to provide information to users who are blind. For this reason, we have excluded text based feedback from our design considerations in Table 3.1.

The remaining ten participants reported that they did not use assistive devices. One participant did not use assistive devices as there were not any available when they were younger, two did not know what assistive devices are available, four reported not needing assistive devices, and
one participant noted that they do not use one as they only go places that they are familiar with. These participants wanted devices that could scan barcodes, direct them to a specific product or area in a store (4), and read labels or other text (8) including information to identify the product, allergy related information, prices, etc. These results further support the need to provide assistance in navigating around stores to find desired products and reading text found on products or store signs.

When asked how they would want the device to convey information to them, four wanted devices to communicate through large text on a screen where the text size can be adjusted, four participants prefer voice feedback methods, and two wanted both voice and text options. One participant reported that they do not want a device that speaks to them in public.

Then we wanted to identify what type of device participants would prefer. Eleven participants wanted to be able to hold their device, four wanted to wear their device, and one selected “other” noting that both have downsides. Participants reported that they chose their device type to increase attention (2), have a hands free assistance (3), for ease of use, specifically when holding objects (2), so that their vision isn’t blocked and nothing is on their head, and to be able to put it in their pocket when they are not using it (4).

To shop independently, participants wanted assistance with reading signs, audible descriptions of items, assistance to avoid collisions while walking, navigating and understanding the layout of a new store, accessible card readers and checkout aisles, a machine in stores that says where items are, and help from employees.

Ultimately from this questionnaire we have identified a set of design considerations, as seen through the bold responses in Table 3.1. Additionally, we found that there is a need to provide assistance in navigating around unfamiliar stores to find desired products. This finding was also supported by Bhowmick and Hazarika, who discussed the importance of providing navigation assistance in unknown indoor environments [9]. From the formative questionnaire, we also found that people who are B/LV wanted help reading text found on products or store signs to enable them
Table 3.1: Design Considerations from Formative Survey

<table>
<thead>
<tr>
<th>Design Consideration</th>
<th>Participants’ Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>What challenges do you face when shopping at a new store?</td>
<td><strong>Finding desired products (9/16)</strong>&lt;br&gt;<strong>Not knowing the floor plan or layout of the store (9/16)</strong>&lt;br&gt;<strong>Trouble with labels and text on products or store signs (3/16)</strong>&lt;br&gt;<strong>Avoiding obstacles (2/16)</strong>&lt;br&gt;<strong>Checkout (1/16)</strong></td>
</tr>
<tr>
<td>How does your device communicate information to you? (Only participants who have used assistive devices before)</td>
<td><strong>Voice or speech (3/6)</strong>&lt;br&gt;<strong>Text (1/6)</strong>&lt;br&gt;<strong>Magnification (2/6)</strong></td>
</tr>
<tr>
<td>How would you want a device to communicate this information to you? (Only participants who have not used assistive devices before)</td>
<td><strong>Voice or speech (4/10)</strong>&lt;br&gt;<strong>Large text on a screen (4/10)</strong>&lt;br&gt;<strong>Both voice and text (2/10)</strong></td>
</tr>
<tr>
<td>What features would an assistive device need to help you? AND What kind of information would you want an assistive device to communicate to you? (Only participants who have not used assistive devices before)</td>
<td><strong>Read labels or other text (8/10)</strong>&lt;br&gt;<strong>Finding desired products or locations (4/10)</strong>&lt;br&gt;<strong>Scan barcodes (1/10)</strong>&lt;br&gt;<strong>Help with checkout pin pads (1/10)</strong>&lt;br&gt;<strong>Avoiding obstacles (1/10)</strong></td>
</tr>
<tr>
<td>Would you rather use an assistive device that you hold in your hands?</td>
<td><strong>Hold in your hands (11/16)</strong>&lt;br&gt;<strong>Wear (4/16)</strong>&lt;br&gt;<strong>Other (1/16)</strong></td>
</tr>
</tbody>
</table>

to shop independently in new stores. Similarly, Brady et al. found that many people who are B/LV would request assistance in their everyday lives to identify objects, read text, and for descriptions of objects [10]. We limit the scope of our work to provide shopping assistance in grocery stores as grocery stores were the most popular type of store at which our participants reported shopping.

### 3.3 Sensor Packages

Some stores, such as home improvement stores, have departments located in both indoor and outdoor settings. When designing an assistive device to provide shopping assistance, it is important
to consider including a sensor package that would enable a device to work in both indoor and outdoor environments. We expanded upon the initial survey [49], by investigating the sensor packages used in indoor and outdoor environments. From the 101 devices in the expanded survey, 16 were evaluated in both indoor and outdoor environments. Table 3.2 breaks down the sensor packages used in both indoor and outdoor environments.

Table 3.2: Sensor packages used by 16 assistive devices that were evaluated in both indoor and outdoor environments

<table>
<thead>
<tr>
<th>Sensor Package</th>
<th>[50]</th>
<th>[51]</th>
<th>[22]</th>
<th>[52]</th>
<th>[53]</th>
<th>[54]</th>
<th>[55]</th>
<th>[56]</th>
<th>[57]</th>
<th>[58]</th>
<th>[36]</th>
<th>[59]</th>
<th>[60]</th>
<th>[61]</th>
<th>[62]</th>
<th>[63]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereo Camera</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3d Time of Flight Camera</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unspecified RGBD</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonar</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motion Sensor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFID Reader</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetometer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gyroscope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLAM/Maps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound Beacons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accelerometer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bluetooth Beacons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The devices in this table can be further analyzed by the assistance that they provide. Four devices focused on providing navigation instructions [50, 51, 36, 59]. Of these four devices, two required the environment to be modified to include beacons [50, 59]. Rather than modifying the environment, sensor packages including sensors for object detection and to estimate the user’s orientation and location have been used [51]. Additionally, Poggi and Mattoccia proposed a device that uses sensors for object detection as well as GPS to estimate the users location and orientation through Google Maps [36]. Two of the devices evaluated in indoor and outdoor environments provided navigation as well as information gathering assistance [55, 63]. Both of these devices used sensor packages including sensors for object detection and to estimate the user’s orientation and location.

Modern day smartphones are very lightweight and compact devices that typically include an
RGBD camera that can be used for object detection, have GPS for outdoor positioning, and have an internal IMU that can be used to estimate changes in position as well as the user’s orientation. Using a smartphone and its internal sensors is also convenient for people who are B/LV, as they can use a multipurpose device, which many people who are B/LV already own [64], rather than requiring them to purchase or carry around a device specifically designed for shopping assistance.

3.4 Proposed Device Feedback Questionnaire

After the formative survey and conducting an analysis of sensor packages, we began to think about how to design a smartphone application to assist people who are B/LV. Ultimately, we needed to ask follow-up questions regarding aspects of the specific design, to understand the target users’ preferences on these design decisions. The follow-up survey was conducted online on Qualtrics and was approved by the university’s IRB. To recruit participants, we reached out to the mailing list of participants that we compiled from the first survey, based on an indication that they were willing to be contacted for future surveys; this list was comprised of 11 participants. A total of five of the 11 participants participated in this survey, as well as a new participant, ultimately resulting in six responses to this survey. This survey was expected to take an hour at most. As a compensation for the participants’ time and effort we sent our $15 Amazon gift cards to everyone who completed the survey. The full survey is in Appendix B.

The participants’ ages were in the range of 18 to 51 (\( M = 30.8, \text{SD} = 14.20 \)). Four participants identify as female and two as male. Half of the participants had their vision impairment since birth, one since they were 2, one since they were 3, and one since they were 34. A distribution of our participants’ degrees of visual impairment can be seen in Figure 3.1.

All participants thought that a smartphone app that uses a map of the store to help navigate to the products they are looking for would be beneficial. The feedback methods that participants preferred for this device were not consistent, showing the necessity for customization. One par-
Figure 3.2: Distribution of reported degree of visual impairment across participants who participated in our proposed device feedback questionnaire.

Based on the results of our formative survey, most participants seemed to prefer speech feedback methods. One participant noted that they did not want the device to speak aloud in public, thus we wanted to explore the option of using bluetooth headphones to maintain the user’s privacy. When asked about using a bluetooth earpiece, all participants said that they would be willing to use one, but one participant noted that they want would only use it if it did not interfere with hearing what was going on around them.

In the formative questionnaire, many participants indicated that they wanted assistance navigating the layout of new stores. This would require a map of the store; to keep the maps up to date we have considered using community updated maps. When asked if stores should provide the maps of their stores, four participants strongly agreed and one agreed. Additionally, five partici-
pants answered that they were willing to contribute data to update the maps. The sixth participant
did not respond to either of these questions.

Five participants indicated that an app which both brings them to their desired product and
is capable of reading store signs and labels would be beneficial to them. While scanning labels,
these five participants responded that they were willing to contribute their data to the map. Two
participants wanted this app to read all text visible to the camera, one wanted it to read information
that they ask for, and three wanted both options.

Participants did have some privacy concerns with this device. They wanted to ensure that
their data was not shared with third parties. Additionally they wanted to ensure that any private
information, such as credit card information that might have been scanned, is not shared or saved.

From this questionnaire, we have identified that our planned device that assists users to their
desired product and is capable of reading store signs and labels would be beneficial to people who
are B/LV. We also found that participants would be willing to use a Bluetooth earpiece with this
device. Additionally, our participants supported the design decision of community updated maps,
as long as privacy concerns were addressed.

### 3.5 Summary

From the responses and feedback in our surveys, we have determined that people who are B/LV
would benefit from a device that provides shopping assistance including navigating to their desired
product and reading text found on product labels or store signs. We focused on providing shopping
assistance at grocery stores since grocery stores were the most popular type of store our participants
shopped at. Users mostly wanted their device to use speech feedback, and participants indicated
that they would be willing to use Bluetooth headphones. We have chosen to design a smartphone
application as many people who are B/LV already own smartphones [64], thus eliminating the need
for users to purchase and carry an additional device. Additionally, modern smartphones contain
all sensors necessary for indoor and outdoor sensing. Additionally, the application will read and contribute to a shared map, which is updated online as people use the app to ensure that the maps are kept as up to date as possible.
Chapter 4

System Implementation

This chapter first gives an overview of the designed Android based application and its functions. Then we discuss the process for selecting a smartphone for this application. After we discuss user input to the application. Finally we discuss the localization, label parsing, product identification, mapping, and route finding modules.

4.1 System Overview

Figure 4.1 shows the application process flow. The application uses camera images, RGBD point clouds, and accelerometer, gyroscope, and gps sensor data. A combination of GPS, camera, and RGBD data is used to localize the device in the store. Then the application receives the latest community updated map from a database, and loads the map based on its localization. Next the app begins scanning and adding or updating the locations of new products and walkable paths to the shared map. When a user presses anywhere on the screen and uses a valid voice command, the application will then enter the corresponding mode.

If the user wants to find a product, the application will generate a route to that product and will provide navigation instructions to the user through speech. In this mode, the app continues to scan products on the shelves as the user walks to their desired product.
Figure 4.1: A flowchart diagram showing how the app operates. The product identifier and map update modules loop until the application ends. The user can interrupt these modules by pressing down on the screen, and then they enter assistive mode. In assistive modes that use the label reader, the product identifier and map updates are disabled. In assistive modes that find products, the A* path planning module generates a path to the desired product and then the application guides the user through the path while also starting the product identification and map update loop.

If the user wants to get information from the product label, such as the number of calories, then the application temporarily disables map updates and processes the product label, providing the desired information to the user through speech.

These modules will be described in more detail below.

### 4.2 Smartphone Selection

When selecting a smartphone for this application in May 2020, we had several initial search criteria. First, we limited our search to Android phones. This decision was made to make development easier; however, it is worth noting that most people who are B/LV have Apple iPhones [64] so this
application will need to be ported in the future. In order to detect objects that a user scans, we need good cameras and access to depth data. Thus, we further limited our search to ARCore supported devices that have access to the Depth API. Google lists these supported devices on their website\(^1\).

Next, we recorded other specs that Google had provided in their list, the release year, cost, processor, battery life (mAh), if it had an audio jack, if it was compatible with Bluetooth, and if it had accessibility options. We found that all Android smartphones come with Talkback for accessibility, therefore accessibility was no longer a limiting criteria. We then identified cost, processor, and battery life as the most important factors to consider in the device. Then we compared all devices across these three categories and narrowed our selection down to two phone models: Google Pixel and Samsung Galaxy. Next, we decided to only consider devices that were being sold new, to reduce any inconveniences of purchasing a damaged product. The Samsung Galaxy S20 was being sold new for around the same price as the S21 and the Google Pixel 5, thus we excluded it since it was an older model. From the remaining two devices, the Google Pixel has a better camera by 4MPs, but it has a dated processor. Thus, we decided to select the Samsung Galaxy S21 phone, as the processor was better and the camera was decent. Figure 4.2 shows the phone used in this work. A better processor was selected as this application was expected to do quite a significant amount of on board processing for object identification and for SLAM mapping. Finally, when selecting a device, we decided to choose an unlocked or GSM phone to enable more coverage if this device were to be shipped outside of the United States. A link to our spreadsheet comparing devices can be found in the footnote\(^2\).

### 4.3 User Input Commands

To enable users to request assistance using our application, we have designed a set of verbal commands. Users can press and hold anywhere on the phone screen, and then say the command

\(^{1}\)https://developers.google.com/ar/devices#google_play

\(^{2}\)https://docs.google.com/spreadsheets/d/1f5Em0p8Ww1CjRdU5Mx13SJOQ38viCJuHnV_KYgGgbfo/edit?usp=sharing
corresponding to the assistance they are looking for. These commands offer assistance finding desired products and retrieving information from product labels or signs. A table describing these commands at a high level can be seen in Table 4.1.

The speech recognition module was built on top of the Android SpeechRecognizer\(^3\). The SpeechRecognizer generates a string based on the user’s spoken input. This string is then cleaned up and parsed into a command and request based on the structure shown in Table 4.1. If a command is not recognized, or if the request fails, the user is notified through speech output. Additionally, the user is notified when the “Stop” command succeeds, as in initial testing users were unsure what mode they were in after using that command.

\(^3\)https://developer.android.com/reference/android/speech/SpeechRecognizer
Speech was chosen as a user input method over using Talkback and menu options, as if we used menu options then users would have to type in their desired products or nutrition term. A small spelling mistake could throw off the system, as the system would be unable to distinguish if the product does not exist or if it was a spelling mistake. Thus, we chose speech input, as the speech recognition module is able to accurately determine what the user says, potentially reducing frustration.

One limitation of speech input is that if another customer is talking at the same time as the user is stating a command, the speech recognition module can not distinguish between the two voices, leading to a failed command. Recognizing speech based on voice is out of scope of this project, but can be investigated in future work. Future work could also investigate using natural language processing techniques to solve the issue with spelling mistakes in user input. A comparison of menus using Talkback, speech, or other input methods should be investigated in future work.

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find in store [product name]</td>
<td>Guides the user to the desired product in the store if it exists in the labeled map. Otherwise keeps scanning until it sees the product.</td>
</tr>
<tr>
<td>Find in shelf [product name]</td>
<td>Guides the user to the desired product on the shelf in front of them. Otherwise keeps scanning until it sees the product.</td>
</tr>
<tr>
<td>Scan shelf</td>
<td>As the user moves their hand holding the device across a shelf, it reads the name of the product currently in front of the device.</td>
</tr>
<tr>
<td>Read</td>
<td>Reads all text in front of the camera.</td>
</tr>
<tr>
<td>How many [nutrition term]</td>
<td>Searches the label for the nutrition term. Reports the amount if it is found, otherwise keeps scanning until it finds the term.</td>
</tr>
<tr>
<td>List allergies</td>
<td>Searches the label for allergy information. Reports a list of potential allergies to the product if found, otherwise keeps scanning until it finds allergy information.</td>
</tr>
<tr>
<td>Repeat</td>
<td>Repeats the last phrase it said.</td>
</tr>
<tr>
<td>Stop</td>
<td>Stops the current command.</td>
</tr>
</tbody>
</table>
4.4 ARCore

ARCore and its predecessor Project Tango have been used in several assistive devices for people who are B/LV. ASSIST provides indoor navigation assistance using a combination of bluetooth beacons and a Project Tango device [65]. Project Tango based assistive devices have also been developed to provide assistance with obstacle avoidance [66, 67]. ARCore based devices have additionally been used to provide navigation assistance [68, 35]. ARCore provides out of the box SLAM and localization and is supported on most modern Android smartphones, thus is a very powerful tool for smartphone based assistive devices for people who are B/LV. Additionally, ARCore provides easy access to RGBD data, through its depth API, from the phone’s time of flight or stereo cameras.

The core of our application is built off of ARCore. Behind the scenes, ARCore sessions use SLAM to track the device’s current position relative to where it started. This is done using a combination of the features extracted from the phone’s RGBD camera as well as its internal IMU sensor.

The world locations of points update as the application scans more of its environment, thus we need to use anchors to represent these dynamically updated positions. Anchors enable us to link a pose to a location in the world, which will be updated as the system learns more about the world and updates the map accordingly. This is important as our map needs to store positional information on where products were last known to be in the store layout. These anchors can also be converted into cloud anchors and shared between applications.

Cloud anchors are anchors that are tied to a set of features. These anchors must be resolved before they can be used by scanning the environment until a sufficiently similar set of features is found. One limitation of using cloud anchors is that only 20 cloud anchors can be resolved at a time. Thus using cloud anchors for all data that we want shared between applications is not feasible.
4.5 Database of Community Updated Maps

To enable multiple users to use and update a common map, we store all necessary information used to recreate the map in an online database. This ensures that the maps of stores are kept as up to date as possible.

We considered using AWS\(^4\) and Firebase\(^5\) for our online database. Ultimately, we decided to go with a realtime Firebase database as it was easier to setup and manage, and, more importantly, was cheaper to host. Firebase uses a pay as you go method for payments, where users only pay for features after they have exhausted the free monthly quotas. Currently, the application has not used more than the monthly quotas. While the application is currently on a small scale, if the scale was larger and the quotas were to be surpassed, the pricing Firebase offers is still very affordable.

An screenshot of database entries for the community map of our testing site can be seen in Figure 4.3. The community maps are indexed using the street address obtained using the device’s GPS location. If no entry exists for a given street address then a new entry is created. When a new entry is created, an ARCore Anchor id for the origin is stored in the database. If an entry for a street address exists, then the origin anchor id, walkable paths, and product locations and names are loaded into the user’s local map. Each walkable path and product database entry has a field indicating the last time it was updated. Entries that were last updated outside of the allowed range are not loaded into the user’s local map and they are removed from the database. This allows the community maps to remove old and potentially inaccurate locations of walkable paths or products in the store. The application adds all newly discovered walkable paths and products to the database as soon as they are discovered. The positions stored in the database are positions relative to the origin of the store. This allows us to instantly resolve all product and walkable path locations as soon as the origin is resolved. If the application discovers a product that already exists in the database, then it updates its last updated date.

\(^4\)https://aws.amazon.com/
\(^5\)https://firebase.google.com/
4.6 Localization

To localize the device in the shared map of the store, first the application gets a street address corresponding to the GPS latitude and longitude. This street address is then used to look up the shared map, in the online database, that corresponds to the store in which the user is currently shopping.

Once the store has been identified, the origin of the store (i.e., the entrance) needs to be resolved. If the origin anchor id entry exists, then the ARCore anchor id for the origin is loaded. Then the application indicates to the user that they should walk around slowly and move their phone side to side until the origin is resolved. Otherwise, the app requests that the user slowly walks around moving their phone side to side until enough features have been detected for the origin to be accurately resolved in the future. Then a new ARCore anchor id for the origin is created and added to the database.

This method for localization requires the user to scan a fairly large area at the entrance of the
store in order to accurately resolve the origin. Alternative solutions can resolve the user’s current location in less time, without requiring the user to scan a large area.

One alternative method to resolve the origin of the store includes using QR codes, which could be attached to doors or walls near the entrance of the store. In preparation for a user study, we needed to reduce the time spent localizing the application. Thus, we implemented a localization method using the QR code scanner, ZXing\(^6\). Once a QR code has been identified, we raycast from its 2D pixel coordinates to find the depth point closest to the QR code. An ARCore anchor representing the origin is then generated at the world location of the depth point. It is not ideal to require stores to add QR codes at their entrances, however, the two localization methods we include are sufficient for the scope of this project.

Assistive devices that provide navigation assistance have also localized using bluetooth beacons [65, 69, 50] and RFID [33, 70, 63], however, much like QR code based localization, these methods requires modifications to stores in order to function.

Future work can implement WiFi based localization in this application, such as Liu et al.’s approach using peer assisted localization [71]. This approach would not require the user to scan their environment at all to resolve the origin, allowing them to immediately begin shopping using the app.

Once the origin has been resolved, all walkable paths and products are loaded into the user’s local map. These positions are saved in the database as positions relative to the origin, therefore their positions will all be updated relative to the pose of the origin when they are added into the user’s local map.

### 4.7 Label Parser

To extract information such as the product name, nutrition information, or allergy related information we have developed a label parser. This label parser is built using Google’s ML Kit Vision API.

\(^6\)https://github.com/zxing/zxing

32
The label parser takes in an image from the device. The higher the image resolution, the better the character recognition quality will be, but this comes with a performance trade-off. Then the label parser performs optical character recognition (OCR) on the image, resulting in a set of text blocks. These text blocks can then be divided into lines of text, which can be further divided into elements.

These text blocks are then parsed to extract the information desired by the user. To extract the desired information, we have designed a set of rules based on patterns that we noticed across products. These rules can be found in the subsections below. While these rules do not encompass every scenario found in every product label, as we will discuss in more detail in Section 5, due to the scope of this project we chose to implement a rule based system rather than implementing natural language processing techniques.

If the bounding box of the desired information is touching the border of the image, then the parser does not extract the desired information as it is likely cut off. The ability to provide partial information could be added in future work.

4.7.1 Product Name

The product name, brand name, and any product descriptors necessary to differentiate similar products (e.g., “Zero Sugar”) are usually the largest text on the front of a product. With the assumption that the product is facing outwards from the shelf, we define a rule to search the label for the largest text. To extract the product name, the label parser first searches each element in every text block and finds the element with the largest bounding box height. All elements within a tolerance of the maximum bounding box height are also added to the resulting product name. Eligible elements are added based on the order they appear from left to right and top to bottom. Examples of product labels can be seen in Figure 4.4.
4.7.2 Nutrition Information

Nutrition information, such as the number of calories, total fat, amount of sodium, total sugar, etc. can be found on the nutrition facts label, which is follows a fairly standardized format\(^7\).

With this standardized format, we are able to make assumptions that all nutrition information will be presented in the same format where the nutrition term is followed by an amount which is then followed by a percentage of daily value. Thus, to extract the desired nutrition information, we first search the label for the nutrition term. If the term is found, and is followed by a number, we then extract the number. Examples of nutrition facts labels can be seen in Figure 4.5.

4.7.3 Allergy Information

Allergy information is found under the nutrition facts label and is indicated by the word “Contains” which in some cases is then followed by a colon. To extract the allergy related information from a label, the label parser searches the label for the word “contains” and removes all instances of the

\(^7\)https://www.fda.gov/media/99151/download
colon character. The word “contains” is not uniquely used in the context of allergy information. If the word “contains” is followed by a digit then it is skipped, an example of a label where this would be the case can be seen in the label on the right in Figure 4.6. If it is followed by a string then it keeps reading lines until it finds a period. While performing this search, any element that has a bounding box located to the left of the bounding box of the word “contains” is skipped. This removes undesired words in multi-column labels. If the label parser finds a period then all text fitting the above criteria is extracted as allergy information. If it does not find a period, then only the text immediately following the initial word “contains” is extracted. Examples of allergy information on labels can be seen at the bottom of both labels in Figure 4.6.
4.8 Product Identification

Product identification is a hard problem to solve as creating a model that can identify all products offered in a store would require large amounts of labeled training data which would be costly to collect. This amount of labeled data would only increase as new products are added to the shelves. Additionally, special promotions or rebranding can also result in lower performance depending on how much the labels change compared to the training data. To create a robust product identifier, without requiring labeled training data for each product we would like to identify, we create a product identification module that uses a machine learning model in combination with our label reader.

4.8.1 Dataset

Several datasets for grocery store products exist. Many of these datasets are small, specialized to train on very specific products [72, 73]. One large and generalizable grocery store product dataset is the SKU-110K dataset [8] which includes training data of 110,712 classes of products found on densely packed retail shelves around the world. These 110,712 classes of objects are all classified as an “object” rather than assigning individual product names to each class. Thus, when a machine learning model is trained on this data, it will learn representations of what defines a generic product.

4.8.2 Object Detection Model

To identify the bounding box associated with products on shelves we trained a small YOLO-v5 [7] model (YOLO-v5s) on the SKU-110K dataset [8]. YOLO-v5 is a state of the art one-stage object detector [74]. We selected a one-stage object detector rather than a two-stage object detector as faster recognition speeds are more important to our application than accuracy. This is because we only need a rough bounding box around products to pass to our label parser. To integrate this model
into an Android application, we then export it to a TFLite model. Using this model, we input an image from the device’s camera and then the model outputs the pixel coordinates, width, height, and confidence for each predicted product. Any detection with a confidence score below 60% is rejected. The pixel coordinates, width, and height can be used to generate a bounding box for each detection. An example of product detection results can be seen in Figure 4.7, where there are multiple overlapping detections for each product as seen by the unaligned squares on the products.

Figure 4.7: Example product detection results from our product detection model. Note, the bounding box was drawn four times thinner in this case compared to all other figures. We did so to make the overlapping detections more noticeable.
4.8.3 Non-Maximum Suppression

The model may predict more than one bounding box for each object as seen in Figure 4.7. Therefore we must apply a non-maximum suppression (NMS) filter to find the most appropriate bounding box for each object. To do so, we first select the prediction with the highest confidence score. Then we compare it to the intersection over union (IOU) of each other prediction. If the IOU is greater than a threshold, \( \text{IOU} \text{.\textit{THRESHOLD}} = 0.5 \) in our implementation, then the prediction overlaps too much with the prediction that currently has the highest confidence score and can be removed. We then repeat this process until every prediction has been analyzed. An example of the resulting detection results can be seen in Figure 4.8. In this figure, you can see that there is only one detection result for each product after applying this filter, unlike in Figure 4.7.

4.8.4 Product Identification

Once we have received the bounding boxes corresponding to the detected products, we can then classify each product by product name using our label parser. First, we extract the image data within the bounding box of the detected product. Then we pass this image data to our label parser, which then performs a product name extraction. The resulting product name is used to label the product.

4.8.5 Mapping Products

Once we have the detected product location in pixel coordinates and the associated product label, we can then add the product to the community map. We use ARCore anchors to maintain the relative positions of these products as the map is updated. To create an anchor we raycast from its 2D pixel coordinates of the detection to find the depth point closest to the target location. Figure 4.10 shows examples of the resulting anchor in two challenge cases, while the user was walking (left) and through the doors in a freezer aisle (right). This gives us the 3D world position
Figure 4.8: Example product detection results from our product detection model after applying our NMS filter.

of the product. Then we calculate the translation and rotation offset of the detected product from the origin. Finally, we submit a new product entry to the Firebase database under the user’s current street address. This entry is composed of the product’s name, offset from the origin, and the date that it was created or modified as seen in Figure 4.9.
4.9 Navigation

When users use the *find* command to request directions to their desired product, our navigation module uses its internal representation of the map and the device’s world position to generate a path to the product. Our application uses ARCore for SLAM. ARCore does not provide a means of obtaining the user’s local map, therefore, we have to create our own representation of the map. This project was not focused on navigation, as our initial survey [49] found that a large percentage of other devices have explored this subfield previously. Thus, we have rather provided a simple navigation implementation to enable us to focus on the novel modules of our device. One popular navigation stack in robotics is Move Base\(^8\). Move Base uses the sensor’s world pose and the sensor’s data to generate and update a map which can be used for its path planning algorithm. This approach has previously been used with Project Tango devices [75], showing that future work can apply this method using ARCore as well. Future work should also integrate dynamic obstacle avoidance, such as that used to provide navigation assistance [36], aerial obstacle avoidance [61], and avoiding collisions with people [30].

\(^8\)http://wiki.ros.org/move_base
Figure 4.10: Anchor generated at the detected product location. The products on the left were detected while the user was walking. The products on the right were detected through the doors in a freezer aisle.

4.9.1 Walkable Paths

In our simplified navigation implementation, it was necessary to ensure that the user is only guided through areas that are not obstructed by shelves, checkout counters, or walls. To do so, we record the position of the user’s device as a walkable waypoint as the user walks through the store. A new waypoint is created if a user walks to an area that is more than our offset of 0.5 meters away from another waypoint. We found that a value of 0.5 provided sufficient numbers of waypoints to assist a user with navigating around aisles.

4.9.2 Map Representation

To represent our map of walkable paths in a store, we use a weighted undirected graph. When a new walkable path waypoint is created, an edge is added from the last waypoint the user had visited to the new waypoint. The weight associated with this edge is the Euclidean distance between the two waypoints. Additionally, our system loops through every walkable path waypoint and creates a new edge between all waypoints that are within a predefined tolerance of 0.9 meters. This value
was selected as it connected all adjacent waypoints, and did not connect waypoints that were on the other side of aisles.

4.9.3 Product Path Finding

When a user uses the *find* command to request directions to a particular product, our system searches all products to find one with a matching name. Once this product is found, the graph is then searched to find the waypoint closest to the desired product. Then our navigation module uses an A* search to find the shortest path from the user’s current position to their desired product.

4.10 Feedback Methods

As noted in our formative questionnaires, participants wanted assistance to be provided through speech. The Android TextToSpeech module \(^9\) was used to provide assistance through speech. Additionally, we took inspiration from applications such as Seeing AI, which uses audio tones to indicate when the device is scanning for a barcode but no barcode is in view of the camera.

4.10.1 Text Reading Assistance

Our device offers three different methods of text reading assistance. The assistance given to the user through speech output is summarized in Table 4.2. In addition to the speech output, if the nutrition term or list of allergies is not found the device plays an audio tone to indicate that the user should rotate the product.

4.10.2 Navigation Assistance

To provide navigation assistance, our device uses speech to indicate the direction that the user should travel. If the user is not facing the next waypoint in the path, the device indicates that they


42
should “Turn Left”, “Turn Right”, or “Turn Around”. The “Turn Around” command is broken into three steps, first the device stated “Turn Around”, then it repeats “Keep Turning” until the user has turned to face the next waypoint, and finally it states “Stop Turning” to indicate that the user should stop turning around. If the user is facing the next waypoint, the device will state “Walk Forward”. Through initial piloting of this device, we found that these commands worked better than specifying a distance that the user should turn or walk forward.

### 4.10.3 Assistance Reaching to the Product

Several groups have investigated different methods for providing directional guidance for a person who is B/LV. One directional guidance device uses sonification to communicate the direction of an object that the user desired [55]. Wristbands containing vibrotactile arrays have also been used to communicate the direction that a user should move their hand [76]. Kim et al. developed a directional guidance device for interacting with wall mounted user interfaces [77]. This device was a smartphone that was modified to include four vibration locations. These vibration locations was used to communicate to the user which direction they should move their hand. One limitation of this approach is that it required the smartphone to be modified to include the four vibration locations. Using an off the shelf smartphone is more convenient as this does not require a user to purchase and install another component onto their phone.

Once the user has arrived at the location in front of their desired product, the device will then indicate how the user should move their hand to position their hand in front of the desired product.
We have chosen to use an audio based feedback method to communicate this information as they can be implemented on a smartphone without any additional hardware. Speech was chosen over sonification as our participants indicated that they would like speech feedback in our formative surveys. Additionally, since we use tones to indicate that the device could not find information, using tone based communication methods to indicate a target position could be confusing and contradictory. Vázquez and Steinfeld found that users preferred speech feedback where the device would repeat the phrases: “up”, “down”, “left”, or “right” to communicate the direction that the user should move their hand to aim a camera [78]. Similarly, our device indicates the direction a user should move their arm by using the following phrases “Move your hand to the left”, “Move your hand to the right”, “Move your hand up”, or “Move your hand down”. Through piloting we also tested specifying distances in the phrases, but this was often confusing to the users.

4.11 Summary

We have designed an Android application to provide shopping assistance to people who are B/LV. This application allows people who are B/LV to use verbal commands to activate different assistance modes. Our application uses ARCore, community updated maps, and our label parser and product identification modules to provide shopping assistance. Next we will describe an initial evaluation of our proposed system.
Chapter 5

Results and Future Work

To evaluate the performance of our application, we have collected a total of 77 pictures of store shelves and a variety of grocery products from two different grocery stores. Out of these pictures, 16 were excluded due to motion blur or excessive light reflection, leaving us with a total of 61 images of grocery store shelves. These pictures include a variety of product types, viewpoints, and other challenging scenarios. We have then run the appropriate modules of our application using the saved images as input. We then analyze scenarios where our device works well, in addition to scenarios where our device does not perform well. Additionally, we discuss our planned user study and future work that can help improve this application.

5.1 Product Detection

Overall, the YOLO-v5 object detection model trained on the SKU-100K dataset performed very well on the images we have collected. This model was able to detect almost all of the products in the front row of the shelf, because products are typically grouped by product type and because the labels of products that are behind other products are occluded, it is only necessary to detect the front row of products in our application. Figure 5.1 shows example product detections on three different shelves containing common product types: canned, boxed, and bottled.
While analyzing product detections on the 61 images of grocery store shelves, we found two types of products that our product detector can not detect well: produce and horizontally laid frozen bagged food. Figure 5.2 shows examples of these scenarios. To enable our product detector to detect these objects, future work could investigate collecting more training data to represent these styles of products.

In addition to these two types of unrecognized products, we also identify a few types of incorrect detections. Our product identifier incorrectly detected reflections of products, products with labels that include other products on their logo, and partially occluded products. Additionally, our product identifier individually detects products that are grouped together by something such as plastic six pack rings. Figure 5.3 shows examples of each of these scenarios.

Through further evaluation of our model’s performance, we found that it does not detect products well when zoomed in to individual or small clusters of products. As seen in Figure 5.4, it does not detect many of the products. This is likely because the training data consisted of further back views of shelves, rather than close examples like this. When users want to scan this close to products in our application, we would recommend that they use the “scan shelf” command to get the product name they are pointing their phone at or the “read” command to read all of the detected
Figure 5.2: Examples of product detections, shown through the red boxes on the bottom center of the shelf, on produce and thin horizontally laid bagged food.

text. Future work can collect more training data for these scenarios, or can look into alternative approaches for product detection.

5.2 Product Identification

While our product name extractor in our prototype is not perfect, it produces recognizable results which shows the feasibility of this approach. Many of the detected results have small inaccuracies in their detected names, however, they are still readable. For example, a brand of canned vegetables, Del Monte, is recognized as “Del monte”, “(Del Monte”, “DeLmonte”, “(Delmont”, “Delmont”, “(Del onte”, “Det monte” etc. Future work can investigate using a dictionary of replacements for slightly misidentified words or a dictionary of grocery related terms and natural language processing techniques to correct words that have small mistakes in them. Additionally,
Figure 5.3: Examples of incorrect product detections, shown through the red boxes, on reflections and products with labels that include other products on their logo, partially occluded products, and products that are grouped together by something such as plastic six pack rings.

future work can investigate using the results of a fuzzy Google search, where the search term is the extracted product name, to identify and replace terms that have small spelling mistakes. Alternatively, future work can investigate using a reverse image search of the product label to identify product names when the device does not get a good reading of the product label.

We have identified that four variables influence the accuracy of our product name recognition including the image resolution of the detected product, the direction the product is facing on the shelf, occlusion of the label, and stylized product names.

5.2.1 Image Resolution

After seeing that many of the results were interpretable, but had small spelling mistakes, we began to investigate how our application performed using images of various resolutions. We investigated four resolutions. First is a downsampled resolution that our device could process fairly quickly. Then using the maximum resolution of the device (2034 x 4032), we took pictures at three distances from the product including viewing the shelf, viewing a cluster of products, or viewing a single product. Examples of these view points can be seen in Figure 5.5.

As seen in Tables 5.1 and 5.2, downsampling the resolution of the image resulted in a much
lower product name detection accuracy. Our initial plan for the app was to perform all processing locally on the device, which is why we originally downsampled the image. The three higher resolution views of the products all resulted in more accurate product name detections compared to the downsampled version. Increasing the resolution would result in more computations, leading to an app that is not running in real time. However, future work could investigate offloading image processing to a server to handle this problem. While the higher resolution images resulted in more accurate product names, there are still small inaccuracies. For example, as seen in Table 5.1 Kellogg’s Frosted Mini Wheats was detected as “Kllggo Frosted Mini Wheag”, “leggs Frosted MI, Wheatg”, and “hellegg? Frosted Mini Whealg” in the higher resolution images. As we mentioned earlier this section, these small errors can potentially be corrected using a dictionary of grocery related terms and natural language processing techniques to correct words that have small mistakes in them.
Figure 5.5: Examples the shelf view (left), cluster of products view (middle), and single product view (right).

5.2.2 Direction Product Faces

When products are not facing with their logo facing forwards, it is very difficult to detect what the product is. This is especially a problem with canned products, as seen in Figure 5.6. In this figure, we can see that the products are on the shelf at various angles, in some cases the logos are partially or not visible at all. Future work can investigate using pattern matching to label them based on the identification of other forward facing products that they are similar to on the shelf. If there is no suitable match, then the product is likely facing backwards and can just be ignored since products are typically grouped with similar products on shelves.

Figure 5.6: Example of canned products on the shelf, where the products’ logos are not all facing forwards.

Additionally, products are sometimes upside-down, on their side, or they might display their product name in a circle, such as the product seen in Figure 5.7. Future work can improve
the system by accounting for the direction that the text is facing when reading the product name. The system can then either transform text that is not printed linearly to make it easier for the text parser, or it can rotate the image if the product is not facing right side up.

![Honey Bunches of Oats](image)

Figure 5.7: Example of a product with text that is written in a circle.

### 5.2.3 Occlusion of Label

Some detected products are also partially occluded. This can either lead to the product name or logo being blocked, as seen in Figure 5.8. If a sign were to occlude the detected product, as we can see on the product on the right in Figure 5.8, then our system can not distinguish between the text that belongs to the product, and the text that belongs to the sign. This ultimately leads to the product potentially being incorrectly identified, in this case our system identified the Large Size Kellogg’s Corn Flakes, as “2% LARGE Kellog CORN FLAKES”. Future work can investigate using pattern matching to identify one product name for a set of similar products on a shelf, where similarity scores are used to determine which portions of each product name are most likely to be correct.
5.2.4 Stylized Product Names

Stylized fonts can also impact the system performance, our system has identified Large Size Kellogg’s Corn Flakes, as “2% LARGE Kellog CORN FLAKES”, “ARGE ellegi CORI FLAKES”, “LARGE SIZE LARGE Ileggi CORI FLAKES”, “LARGE SIZE Mellorg COR FLAKES”, and “LARGES Kollorgs CORN FLAKES”. Then it classified regular Kellogg’s Corn Flakes as “illgg FAKES Klloggs CORN FLAKES”, “Kellugs CORN FLAKES”, “Klloygs CORN FLAKES”, “Kol lengs CORN FLAKES”, “Kollogs CORN FLAKES”, and “Kollougs CORN FLAKES”. All of these are understandable results, this shows that the system has a difficult time parsing the cursive stylized text in which Kellogg’s is written. Additionally, there are products that include graphics as part of their logo. Examples of this can be seen in Figure 5.9, our system detected the left product as “Honey. Nut Cheers”, likely due to the use of a picture of piece of heart-shaped cereal in place of the letter ‘O’ in their product name. The right product was detected as “Post e hs”, also likely due to the stylized ‘O in the product name.
5.3 Label Parsing

Our label parser was capable of extracting nutrition and allergy information from most product labels. The Nutrition Facts label is fairly standardized, however, we noticed a few variations of this label which our rules do not account for.

In the first variation, a nutrition key and its corresponding value are split across multiple lines, as seen in Figure 5.10. While one potential solution would be to check the first element of the next line, that does not solve this particular case. In this case, due to the product size, the object has to be positioned very accurately in order for the camera to be able to see both the key and the value. Even when the camera can see both, since the product is round the text is slightly distorted, making it difficult to interpret the text. Future work could investigate storing the contents of a label and piecing the full label together from the results of multiple camera frames that are accessed as the user rotates the product.
Figure 5.10: Example of a label where nutrition fact key and value are split across two lines.

In the second variation, the key “calories” is written in both English followed by Spanish and then is followed by the value, as seen in Figure 5.11. Future work can investigate using natural language processing techniques for translation to prevent this issue.

Figure 5.11: Example of a label where the calories are labeled with English and Spanish keys.

Finally, some products list their nutrition facts across a table of multiple columns. These columns indicate the nutrition facts for the product contained in the package, as well as for scenarios such as adding milk, adding butter, or for one serving. We found two different types of multi-column nutrition fact labels. Figure 5.12 shows the more common variation while Figure 5.13 shows another variation where the calories are not included in the table, but are still split across two different categories on the right. In both of these scenarios, our device reads the value in the column immediately following the key. Future work should investigate applying other methods to parse these multi-column tables, such as by reading the headers and then conveying both values with their category description.
Figure 5.12: Most common example of a label where nutrition facts are split across multiple columns.

5.4 Store Sign Reading

In our formative questionnaire, some participants indicated that they would like assistance reading store signs. Our device can provide this assistance through the “Read” command, which reads all of the detected text. The left image in Figure 5.14 shows an example of an overhead aisle sign, as well as product grouping signs located on the shelves. The right image in Figure 5.14 shows the resulting detected text which is then read aloud to the user. Future work can improve this assistance by investigating natural language processing techniques to correct for misspellings in the detected text.

5.5 Planned User Study

We designed a user study to gather initial feedback on our prototype device. In particular, we were interested in feedback on how the device communicates to the user and other feedback based on
Figure 5.13: Variation of a label where nutrition facts are split across multiple columns.

how they use the device. Unfortunately, we have not been able to recruit participants who are B/LV for this study at this time, likely due to the pandemic. We have set up mock grocery store aisles using a shelf as seen in Figure 5.15. The shelf contains a variety of grocery store products that will be used in our experiment. Additionally, participants are able to walk around the shelf, as if they were walking into a new aisle in a grocery store. In this user study, participants would complete a variety of tasks that enable them to explore all forms of assistance that our device can provide.

**Label Parsing Assistance**

The first task explores the use of the “How many” command. Participants will be guided to a set of products. The goal is to find the product that matches the description of nutrition information described by the experimenter. For example, participants would be asked to find the product that has the most calories, least total fat, or to find the product that contains $X$ grams of protein.

Next, participants will be guided to a second set of products. This task explores the “List Allergies” command. Participants will be asked to identify products that contain particular allergens, such as peanuts, wheat, milk, eggs, or soy.

Finally, participants will be guided to the last set of products, where they will use the “Scan shelf” command to find the product indicated by the experimenter.
After completing all label parsing related tasks, the participants will respond to a questionnaire including a System Usability Scale [79] and then will be asked to give feedback on how the device can be improved.

**Navigation Assistance**

To gather feedback on the navigation assistance provided by our device, we will load the community map of our experimental set up from our database. This prevents the user from having to initially explore the set up to scan all of the products. This decision is also justified as participants had indicated in our proposed device feedback questionnaire that they believe stores should be providing the maps for our application. It also allows us to manually fine-tune the product names in our labeled map, as in this task we are more interested in feedback on the navigation assistance portion of our application. Additionally, we have disabled community map updates during this experimental evaluation to ensure that all users are using the same maps.

Participants are first guided to the starting zone. The participants’ task is to gather products
on a shopping list read aloud one at a time by the experimenter. The participants will use the “Find in store” command and then the device will guide them to their desired product.

After completing this navigation task, the participants will respond to a questionnaire that asks for feedback on the device and includes a System Usability Scale.

5.6 Additional Future Work

In addition to all of the potential improvements and evaluations described above, there are other improvements that can be made to this device. Currently the application creates one anchor for each detected product. Anchors are updated every time the map is adjusted. It is unnecessary to have more than one of the same object represented by unique anchors. The application should rather identify the region where matching products are and place the marker in the middle of them all.

Another improvement that should be made is to reduce the decay time on products that are detected on the ends of aisles. These products often rotate, and in some cases are used for products
that are on sale. Therefore these products should have a lower decay time so that they are filtered out if not scanned within an appropriate time.

In our initial questionnaires, participants noted that they would like the application to bring them to similar objects if their desired object is not found in the map. To do this we need to categorize products into higher levels of categories. This could be done through using a dictionary of common terms for each category, aisle signs, or shelf signs.

Signs on the shelves could also be used to help identify products. This would help in scenarios where our product detector failed to detect a particular product or if a product does not have a label on it that we can use to identify what it is. This improvement would help identify produce products.

Participants also indicated in our initial questionnaires that they believe stores should provide maps for this application. This is seemingly a viable solution as stores continue to adopt robots into their workforce to check shelf inventory. While the robots take inventory of the shelves, we can leverage maps created or used by these systems as a base line for the community updated maps.

Finally, in our initial questionnaires participants also indicated that they wanted assistance with pinpads and other challenges they face while checking out. Future work could expand upon this application to provide that assistance. Additionally, some participants noted that they also wanted obstacle avoidance assistance while they were walking around the store. Future work can investigate implementing this assistance as well.
Table 5.1: Detected cereal product names across different resolutions.

<table>
<thead>
<tr>
<th>Boxed Product</th>
<th>Detected Product Names</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Downsampled Shelf View Average Resolution: 179 x 80</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View Average Resolution: 538 x 362</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View Average Resolution: 1389 x 995</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Single View Average Resolution: 2656 x 2211</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Kellogg’s Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Kellogg’s Raisin Bran (Sing 2 Promotion)</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> SING TICKET Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> helogc&amp; Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> SD MOVIE TICKET SING Raisin Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> General Mills Raisin Nut Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> Raisin Nut Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> Raisin Nut Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> Raisin Nut Bran Excellent EDIEN</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> Raisin Nut Bran</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Post Honey Bunches of Oats Almonds</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> E OATS ALMOND</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> BUN EY NEY OAI1 ALMONDS</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> BU NEY UNC INCH OATS</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> Post BU A ALMONDS5</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Kellogg’s Frosted Mini Wheats</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> Frosted Mini WInEais</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> Kllggo Frosted Mini Wheag</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> ileggs Frosted MI, Wheatg</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> hellegg? Frosted Mini Whealg</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Market Basket Frosted Shredded Wheat Strawberry</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> MARAT AM FROSTED shredded Wheat</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> FROSTED Shredded Wheat</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> FROSTED Shredded WleaBT</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> FROSTED Shredded Wheat</td>
</tr>
<tr>
<td></td>
<td><strong>Ground Truth:</strong> Kellogg’s Corn Flakes</td>
</tr>
<tr>
<td></td>
<td><strong>Downsampled Shelf View:</strong> CORN FLAKES</td>
</tr>
<tr>
<td></td>
<td><strong>Shelf View:</strong> Kellugs CORN FLAKES</td>
</tr>
<tr>
<td></td>
<td><strong>Cluster View:</strong> Koclogg CORN FLAKEs</td>
</tr>
<tr>
<td></td>
<td><strong>Single View:</strong> Keleygs CORN S FLAKES</td>
</tr>
</tbody>
</table>

60
<table>
<thead>
<tr>
<th>Canned Product</th>
<th>Detected Product Names</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downsampled Shelf View Average Resolution: 69 x 36</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View Average Resolution: 222 x 152</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View Average Resolution: 905 x 606</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Single View Average Resolution: 1952 x 1134</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Lentil</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> NONE DETECTED</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> PROGRESSO LENTIL</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> PROGRESSO</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> PROGRESSO</td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Vegetable</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> UA</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> PROGRESSO</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> ROGRE CLASSIC &quot;ROGRESS VEGETABLE</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> DnaOAS CLASSICS PROGRESSO</td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Hearty Penne</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> AT</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> PROGRESSO HEARTY PENNE</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> VEGETAB PROGRESSO HEARTY PENNE WOARTIELC FLAVON AET 02TL830</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> VEGETABLE CLASSICS PROGRESSU SOU HEARTY PENNE NOANE 02L83</td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Garden Vegetable</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> -</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> PROGRESSO VEGETABLE</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> PROGRESSO0 GARDEN VEGETABLE</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> PROGRESSO</td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Green Split Pea</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> GE</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> VEGETABLE CLASSICS PROGRESSO GREEN SPLIT PEA</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> PROGRESSO GREEN</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> PROGRESSO PR</td>
<td></td>
</tr>
<tr>
<td><strong>Ground Truth:</strong> Progresso Vegetarian Vegetable</td>
<td></td>
</tr>
<tr>
<td><strong>Downsampled Shelf View:</strong> NONE DETECTED</td>
<td></td>
</tr>
<tr>
<td><strong>Shelf View:</strong> VEGETABLE CLASSICS PROGRESSO VEGETARIAN VEGETABLE Tnio</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster View:</strong> CLASSIC GRESSO VEGET TABLE</td>
<td></td>
</tr>
<tr>
<td><strong>Single View:</strong> CLASSIL PROGRESSO EGETARIA VEGETABI</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6

Conclusion

In this work we first conducted a literature search of assistive devices for people who are B/LV and identified a need to help further develop the field by focusing on providing assistance in information gathering and object finding, such as by providing shopping assistance. Then we distributed a formative questionnaire and found that people who are B/LV would benefit from such a device. From this questionnaire, we identified that people who are B/LV wanted assistance navigating around and finding their desired products in unfamiliar stores, as well as assistance reading text based information found on store signs and product labels. We then identified that modern day smartphones both include the necessary sensor packages to provide this assistance and are convenient as many people who are B/LV own smartphones, thus allowing them to use a device they already own rather than requiring them to purchase or carry around a device specifically designed for shopping assistance. Next, we distributed another questionnaire to ask follow-up questions regarding our planned design and to identify the target users’ preferences on design decisions.

Based on the findings of our literature search and the responses to our initial questionnaires, we designed and implemented a prototype device to provide shopping assistance with the goal of enabling people who are B/LV to shop independently in stores they are unfamiliar with. This application uses ARCore to track the user’s current position, the locations of walkable paths, and
the products detected in the store as the user walks around scanning shelves. To keep the maps used by our device up to date, we implemented community updated maps through a Firebase database.

We developed a label parser, using Google’s ML Kit Vision API, that is able to extract various types of text information including product names and descriptors, nutritional information, and allergy information. Ultimately, we found that our label parser was capable of detecting readable product names, with small misspellings. We found that the image resolution of the detected product, the direction the product is facing on the shelf, occlusion of the label, and stylized product names influence the accuracy of the detected product names.

The label parser was successfully able to extract nutritional and allergy information from labels using a set of parsing rules. Future work can investigate further improving the label parser to enable it to work in situations which our rules do not account for, such as labels where keys and values are split across multiple lines, keys written in multiple languages, and multicolumn labels.

For this application, we also developed a product identifier using a YOLO-v5 object detection model trained on the SKU-110K dataset and our label parser. We found that our product identifier was able to detect most products on the shelves. The product identifier did not detect produce and horizontally laid frozen bagged food well, but future work can compensate for that through including more training data. The product identifier worked well, but incorrectly detected reflections of products, products with labels that include other products on their logo, partially occluded products, and products that are grouped together by something such as plastic six pack rings. Additionally, our product identifier does not detect products well when zoomed in to individual or small clusters of products.

Additionally, our device provides assistance navigating to desired products. To generate a route to the user’s desired location, we use an A* search and represent walkable paths as nodes in a weighted undirected graph. Users can use a variety of voice commands to receive assistance from our device. Our device outputs the desired information or instructions through speech.

Ultimately, we have designed a prototype smartphone application that is capable of provid-
ing people who are B/LV with shopping assistance to enable them to shop independently in stores they are unfamiliar with. Our product identifier is robust and enables our device to function in any grocery store environment, without requiring training data specific to the products found in the store or modifications to the store. Additionally, our application builds up a map of the store as users walk around shopping, enabling our application to function in any store, without requiring initial set up by employees. This map is community updated to ensure that it stays as up to date as possible. Additionally, our application enables users to efficiently browse shelves as the application scans products in view of the smartphone’s camera, while the user walks up and down aisles. This prototype smartphone application has shown proof of concept results, which can be further improved in future work.
References


[64] K. Locke, K. Ellis, M. Kent, L. McRae, and G. Peaty, “Smartphones and equal access for people who are blind or have low vision,” Curtin University, 2020.


A Initial Survey

Informed Consent Form

Welcome to the research study! This survey is being conducted by UMass Lowell's Human-Robot Interaction lab under the direction of Professor Holly Yanco to learn about the experiences of people with varying levels of visual impairments.

IRB #: 20-162-YAN-EXM
Approved on: 1/11/2021

If you agree to participate, you will be asked demographic questions and questions regarding your experiences with shopping and assistive technology. Please answer all questions in the order they are presented. This survey should not take more than 45 minutes of your time. Your participation is voluntary and your responses will remain confidential.

With any type of research participation, there is always a risk for stress or disclosure. You are free to skip any questions and can stop the survey at any time. Your decision to discontinue participation in the study will have
no negative consequences. Participation in this study has no direct benefit to you.

As a thank you for your participation, we will provide you a $15 Amazon gift card through email within 3 days after the survey is completed. Your email will be removed after you receive payment, unless if you choose to be contacted for future surveys or device testing. You will need to complete the entire survey in order to receive the incentive. If you withdraw early, you will not receive the incentive.

All survey responses will be kept on password protected computers under the control of the researcher at all times. We do plan to publish and/or disseminate the study results but no identifying information will be used about you and the results will be compiled from multiple participants. If you accidentally provide any identifying information to us, other than your email address, we will remove it from the responses to protect your privacy and confidentiality.
Contact information: For questions about the research, complaints, or problems: Contact Gregory LeMasurier at gregory_lemasurier@student.uml.edu or Holly Yanco at Holly_Yanco@uml.edu.

For questions about your rights as a research participant, complaints, or problems: Contact the UMass Lowell IRB (Institutional Review Board) at 978-934-4134 or at IRB@uml.edu Please select your choice below. You may print a copy of this consent form for your records.

Clicking on the "Agree" button indicates that:
You have read and understand the above information
You voluntarily agree to participate in the research
You are 18 years of age or older
You have a visual impairment, that is not fixable with glasses

☐ Agree
☐ Disagree

Demographics

https://umasslowell.co1.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=SV_8pQFyBErhDlemXz...
Q2. What is the degree of your visual impairment?

- Total blindness
- Profound low vision
- Severe low vision
- Moderate low vision
- None

Q3. How old are you?

- Male
- Female
- Non-binary
- Prefer to self-describe
- Prefer not to answer

Q4. What is your gender identity?

- Male
- Female
- Non-binary
- Prefer to self-describe
- Prefer not to answer

Q4.1. Prefer to self describe (gender identity):
Q5. How old were you when your visual impairment started?

Q6. Before the pandemic, how often would you go out shopping?

- More than once a week
- Once a week
- Once every couple of weeks
- Once a month
- Less frequently than once a month
- Never

Q7. Before the pandemic, what kind of stores did you typically shop at?
Q8. When you shop, do you shop with others?

- Alone
- With family
- With friends
- With a helper
- With support from shop assistants
- I don't shop

Q9. What challenges do you face when shopping at a new store?

Q10. How do you tell similar objects apart, such as two cans of food?
Q11. Finding the item that I want in a new store is

- Very easy
- Easy
- Neither easy nor difficult
- Difficult
- Very difficult

Q12. Do you use an assistive device to help find and identify objects?

- Yes
- No

Answered Yes - Questionnaire

Q12.1. What device do you use?

Q12.2. Please rate your agreement with the following statement. My device is easy to use.

- Very easy
- Easy
- Neither easy nor difficult
- Difficult
- Very difficult
Q12.3. What are some limitations of the device you use?

Q12.4. What do you like about the device you use?

Q12.5. How does your device communicate information to you?
Q12.6. What kind of information does it give you?

Q12.7. Please list any other assistive devices that you have heard of or have used.

Answered No – Questionnaire

Q12.1. Why don’t you use an assistive device?

Q12.2. What features would an assistive device need to help you?
Q12.3. What kind of information would you want an assistive device to communicate to you?

Q12.4. How would you want a device to communicate this information to you?

Q12.5. Please list any assistive devices that you have heard of.

QuestionnaireContinued

Q13. Would you rather use an assistive device that you

- hold in your hands
- wear
Q14. Why do you prefer this type of device?

Q15. What would help you to shop independently?

Q16. Please rate your agreement to the following statement. When receiving items from online shopping, I often face challenges identifying objects.

- Strongly Agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree
Q17. When you drop an object on the floor, how do you go about picking it up?

[Blank]

Q18. If you had an assistive device that would guide your hand to the dropped object would you use it?

- Yes
- Maybe
- No

Q19. Do you have any other comments?

[Blank]

Q20. Thank you! Please enter your email address so we can send you an Amazon gift card as compensation for your time and participation.

[Blank]
Q21. Sorry to ask again but we want to make sure that we send the Amazon gift card to the right email address. Please enter your email address in again.

Q22. Would you be willing to be contacted to provide feedback for a device or to try a device?

- Yes
- No
B Follow Up Survey

Informed Consent Form

Welcome to our follow up research study! This survey is being conducted by UMass Lowell’s Human-Robot Interaction lab under the direction of Professor Holly Yanco to learn about assistive device preferences of people with varying levels of visual impairments.

IRB #: 20-162-YAN-EXM
Approved on: 4/2/2021

If you agree to participate, you will be asked demographic questions and questions regarding your preferences for assistive devices to help you shop. Please answer all questions in the order they are presented. This survey should not take more than 45 minutes of your time. Your participation is voluntary and your responses will remain confidential.

With any type of research participation, there is always a risk for stress or disclosure. You are free to skip any questions and can stop the survey at any time. Your decision to discontinue participation in the study will have
no negative consequences. Participation in this study has no direct benefit to you.

As a thank you for your participation, we will provide you a $15 Amazon gift card through email within 3 days after the survey is completed. Your email will be removed after you receive payment, unless if you choose to be contacted for future surveys or device testing. You will need to complete the entire survey in order to receive the incentive. If you withdraw early, you will not receive the incentive.

All survey responses will be kept on password protected computers under the control of the researcher at all times. We do plan to publish and/or disseminate the study results but no identifying information will be used about you and the results will be compiled from multiple participants. If you accidentally provide any identifying information to us, other than your email address, we will remove it from the responses to protect your privacy and confidentiality.
Contact information: For questions about the research, complaints, or problems: Contact Gregory LeMasurier at gregory_lemasurier@student.uml.edu or Holly Yanco at Holly_Yanco@uml.edu.

For questions about your rights as a research participant, complaints, or problems: Contact the UMass Lowell IRB (Institutional Review Board) at 978-934-4134 or at IRB@uml.edu Please select your choice below. You may print a copy of this consent form for your records.

Clicking on the "Agree" button indicates that:
- You have read and understand the above information
- You voluntarily agree to participate in the research
- You are 18 years of age or older
- You have a visual impairment, that is not fixable with glasses

- Agree
- Disagree
Q1. To track your responses in future surveys we will use a unique id so that we know which responses were belong to each person so that we don't have to ask the same demographics questions in every survey. The unique id is made from the first letter of your last name followed by the first letter of the town you live in followed by the year you were born. For example: ld1998. Please enter your unique id.

As we did not previously track responses using a unique id, we ask that you answer these demographics questions again. Sorry for the inconvenience.

Q2. What is the degree of your visual impairment?

- Total blindness
- Profound low vision
- Severe low vision
- Moderate low vision
- None
Q3. How old are you?

Q4. What is your gender identity?

- Male
- Female
- Non-binary
- Prefer to self-describe
- Prefer not to answer

Q4.1. Prefer to self describe (gender identity):

Q5. How old were you when your visual impairment started?
**General Questions**

Q6. From our initial survey results we found that many participants wanted an assistive device that could help them find products in a store. Would a smartphone app that uses a map of the store to help you navigate to the products you are looking for be beneficial to you?

- Yes
- No

Q7. How would you want an assistive device to communicate instructions to help you navigate to your desired product?

Q8. Would you be willing to use a Bluetooth earpiece as part of an assistive device?

- Yes
- No
Q9. Please rate your agreement with the following statement: Stores should provide the data and maps my device uses to help me find the products I want to purchase

- Strongly Agree
- Agree
- Moderately Agree
- Neutral
- Moderately Disagree
- Disagree
- Strongly Disagree

Q10. While using my assistive device to navigate around the store I would be willing to contribute data from the camera on my phone to help update the map of the store

- Yes
- No

Q11. If you would like to explain your previous answer, please do so here:
Q12. Please rate your agreement with the following statement: If my assistive device is unable to find the product that I want, then I would want my assistive device to bring me to similar products

○ Strongly Agree
○ Agree
○ Moderately Agree
○ Neutral
○ Moderately Disagree
○ Disagree
○ Strongly Disagree

Q13. Do you have any other suggestions for what the device should do if it is unable to find the product that you want?

Q14. We also found that many participants wanted an assistive device to help them read information that is found on product labels and store signs. Would an app that is able to provide this information to you be beneficial?
Q15. Would an app that helps you find products you are looking for and can read product labels or store signs be beneficial to you?

- Yes
- No

Q16. When scanning a product’s label I would be willing to contribute data from the camera on my phone to help update the map of the store

- Yes
- No

Q17. If you would like to explain your previous answer, please do so here:
Q18. The assistive device should read to me:

- All text visible to the camera
- Information that I specifically ask for
- Both

Q19. Would an app that helps you read information on screens or a pin pad be beneficial to you for when you are at a register?

- Yes
- No

Q19.1. How would you want this device to help you?

Q20. Is there any other functionality that you would like in an assistive device to help you shop?


Q21. Do you have any privacy concerns with this device?

Q22. Do you have any additional concerns with this device?

Q23. Do you have any other comments or suggestions regarding this device or this survey?

Thank you

Q24. Thank you! Please enter your email address so we can send you an Amazon gift card as compensation for your time and participation.
Q25. Sorry to ask again but we want to make sure that we send the Amazon gift card to the right email address. Please enter your email address in again.

Q26. Would you be willing to be contacted to provide feedback for a device or to try a device?

- Yes
- No