# Smartphone Based IoT-Controller Framework for Assisting the Blind in Human Robot Interaction

Harish Ram Nambiappan
University of Texas at Arlington
Arlington, Texas, USA
harishram.nambiappan@mavs.uta.edu

Enamul Karim
University of Texas at Arlington
Arlington, Texas, USA

Md Jillur Rahman Saurav
University of Texas at Arlington
Arlington, Texas, USA

Anushka Srivastav
University of Texas at Arlington
Arlington, Texas, USA

Nicholas Gans
University of Texas at Arlington
Research Institute
Fort Worth, Texas, USA

Fillia Makedon
University of Texas at Arlington
Arlington, Texas, USA

## ABSTRACT

In this paper, a novel smartphone-based IoT-Controller Framework is proposed for effective interaction between robots and people who are blind. This framework focuses on assisting visually impaired users in a pick and place task scenario in contrast to previous works, which primarily focus on navigation and localization. The user can give speech commands to the robot using a smartphone application, which is sent to a server for recognition and retrieval of information such as positions and type of object to be grasped. The details are sent to the robot, which performs the commanded task. Preliminary tests with five participants over a total of 20 trials showed that the system had a 85% success rate, and the average time taken for the task to complete was approximately 65 seconds.

## CCS CONCEPTS

• **Human-centered computing** → **Smartphones**; • **Computer systems organization** → **Robotics**.

## KEYWORDS

Internet of Things, Mobile Computing, Human Robot Interaction

**ACM Reference Format:**
Harish Ram Nambiappan, Enamul Karim, Md Jillur Rahman Saurav, Anushka Srivastav, Nicholas Gans, and Fillia Makedon. 2022. Smartphone Based IoT-Controller Framework for Assisting the Blind in Human Robot Interaction. In *The 15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3529190.3534782

## 1 INTRODUCTION

People who are blind or have severe vision impairment face increased difficulty in daily activities such as navigation, identifying and picking up objects, social interaction and communication, etc. Smartphones and mobile devices require unique interfaces for blind users. Users who are blind can use gestures, voice input, external keyboard-based text input and text-to-speech engines for accessing smartphones and mobile devices [3]. Though there are various research works focused on assisting people who are blind, most research currently focuses on navigation and localization. The authors of [4–6] implemented smartphone-based localization and navigation for assisting the blind, while the works in [1, 2, 10] focused on developing a robotic aid to assist people who are blind in navigation. Though there are works such as [8] and [9] that focus on developing human-robot interfaces for assisting people with disabilities, the focus is mostly on directly controlling robotic arm movement through the interface. However, this is of limited help to a user who is blind, as they cannot see the interface or the robotic movements. Therefore, there is a need for an efficient human-robot interactive framework in which a person who is blind can instruct the robot to perform the task, and also can keep the user informed about the progress.

In this paper, a novel smartphone-based Internet of Things (IoT)-controller framework is proposed for assisting people with blindness to engage in human-robot interaction. The framework is implemented in the scenario of locating and picking up objects, then placing them in a location as expected by the user. This is an important issue for people who are blind to live independently, but it has not received significant attention in previous works. The IoT-controller framework interconnects the smartphone, with microphone, with the robotic arm through an intermediate server, which performs speech recognition and information retrieval. The MINA Robot [7], is used to implement this framework with the smartphone application built to be accessed through speech, without using the touch screen, thus making it easy for visually impaired users to use the application.

## 2 PROPOSED SYSTEM

Figure 1 shows the system architecture of the proposed smartphone-based IoT-controller framework, where the smartphone is interconnected with the robotic arm through a server. The MINA Robot consists of a Franka Emika Panda arm mounted on a Summit-XL Steel mobile base and is controlled by a laptop. The user's speech command is recorded by the smartphone application – developed using Android Studio, with the Android Speech Recognizer class using the microphone sensor in the smartphone. The command is sent to the server, where the start and final positions are interpreted


*PETRA '22, June 29-July 1, 2022, Corfu, Greece*

ACM ISBN 978-1-4503-9631-8/22/06…$15.00
https://doi.org/10.1145/3529190.3534782

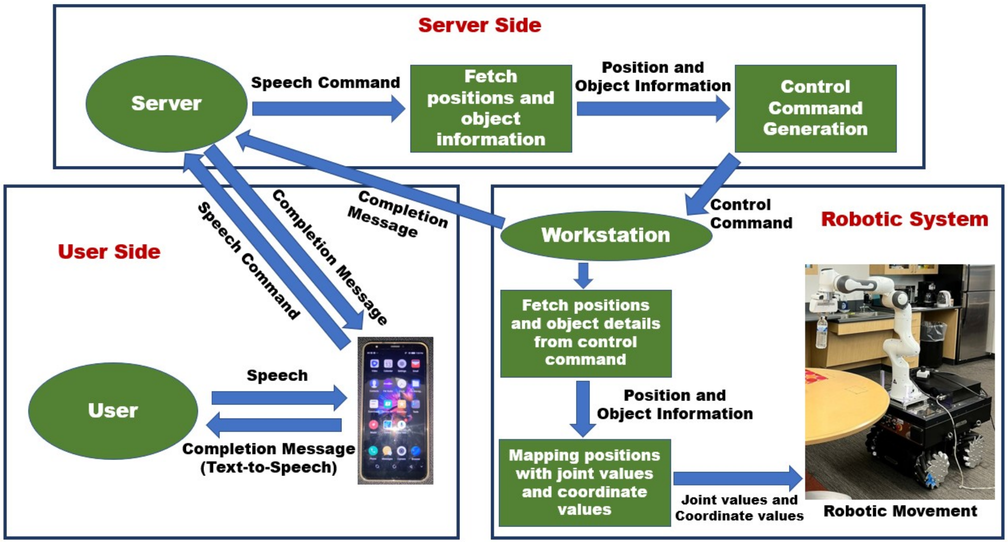


**Figure 1: System Architecture of the IoT-Controller Framework**

from the speech command. The server also contains a database with the position information and the type of object present in that position. The type of object to be grasped at the start position is retrieved from the database in the server. A control command generated with the position and object information is sent to the robotic system workstation. The robotic system workstation interprets the positions and object information from the control command and maps it to the corresponding joint and coordinate values for the robotic movement. Once the robot receives the joint and coordinate values, it performs the pick and place task. On completion of the task, the robotic system sends a task completion message to the server, and the server sends a task completion message to the smartphone application, which notifies the user through text-to-speech. Thus, the IoT-Controller system "connects" the smartphone with the robotic arm through an intermediate server that processes the speech command.

## 3 EXPERIMENTATION

### 3.1 Experimental Methodology

Preliminary experiments were conducted with 5 participants with no visual impairment, and a total of 20 trials were carried out. Five positions, numbered 1 to 5, were marked on a table and three objects - a teabag box, a small bottle, and a big bottle of dish soap were placed at three of those five marked positions. In each trial, the user used the smartphone application by closing their eyes, to replicate a user who is blind, and gave speech commands for the robot to fetch an object from the start position to the end position. All the participants were instructed to say the following commands in the following order: "Connect to System" for connecting to the server; "Move object in position x to position y" where x is the start position and y is the end position. The commands given by the users were not scripted, and users were free to decide which object they want to move from which start to which end position.

Figure 2 shows a user giving the speech command with her closed eyes and the robot picking the object from the start position to the end position. The experimental setup was set in such a way that there were no other objects occupying at the end position where the user wanted to place their object. The main focus was on testing the effectiveness of the current IoT system based on whether the robot correctly moves to the specified start and end positions, the accuracy of the speech recognizer class, and the average time taken to complete the task. The robot was made to move at 10% of its full velocity in order to ensure safety of the robot and the human user interacting with the robot. For time calculation, the time is recorded from the moment the user gives the speech command until the moment the user receives the notification for every successful trial. The success rate and average time taken are calculated.

### 3.2 Experimental Results

Preliminary experiments showed that the system had a 85% success rate, and the average time taken for the robot to perform the task, with 10% of its full velocity, was 65 seconds, therefore showing that the system is effective for users who are blind. The primary reason for the few failures during the trials was errors in speech recognition. The speech recognition process has shown issues in recognizing words with similar phonetics, such as the similarity between words "two" and "to", between "four" and "for" etc. Other minor issues in the speech recognition process include combining two different spoken words in a command, such as recognizing the phrase "position 5 to..." as "position 52" etc. and hence showing errors in the position identification process.

## 4 CONCLUSION AND FUTURE WORK

A smartphone based IoT-controller framework is proposed to assist visually impaired users for effective interaction with robots in human-robot interaction scenario. The user can access a smartphone application through speech, give commands for a pick and

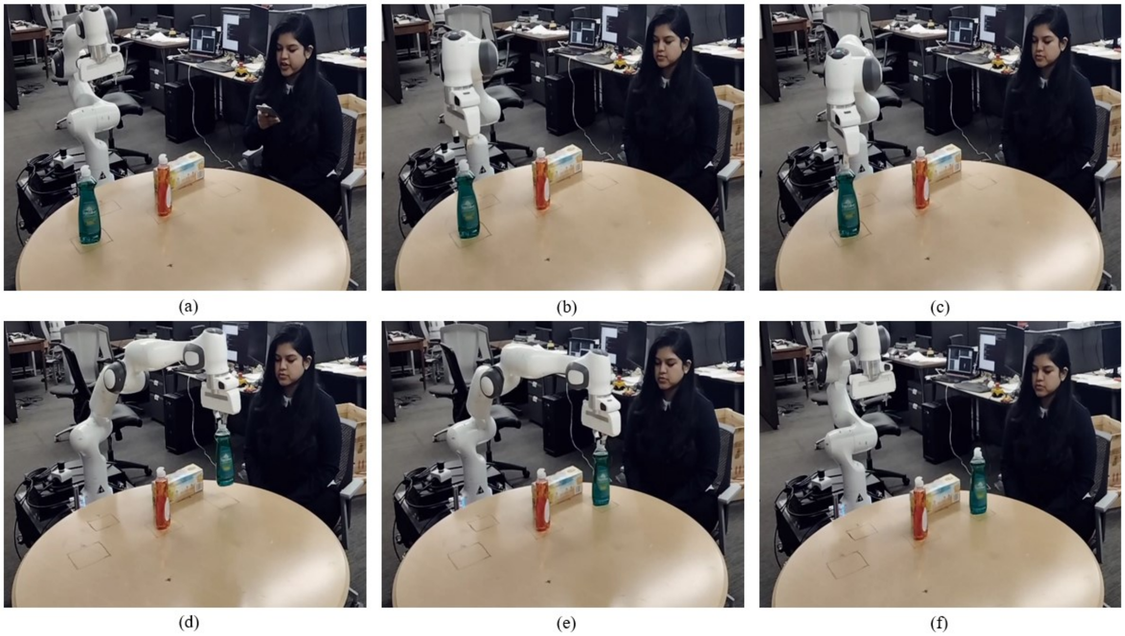

**Figure 2: An example of a user giving a speech command and the robot performing the task. The user's eyes are closed throughout the entire process to replicate the blind scenario. (a) User giving speech command through smartphone application (b) Robot moving to specified start position (c) Robot grasping the object at the start position (d) Robot moving towards the end position (e) Robot placing the object at the end position (f) Robot moving back to neutral pose.**

place task, and the robot performs the task based on the user's speech command. This preliminary work aims to implement the general controller framework and test the integration with the robotic system. This is a basic proof-of-concept and can be extended to various real-life scenarios to assist visually impaired users such as identifying objects, moving objects from one location to a different location in a home or workspace, etc.

There are numerous future improvements and extensions to be done to the system. First, there is a need to add vision-based object detection in the framework to notify the blind users of the object's location and integrate it with robotic motion. To accomplish this task, a camera is to be added to the robot that can perform object detection and send the location of the object to the robot to perform grasping. Second, the Android speech recognizer class records the user's speech only after a buzzer sound and lasts only for a few seconds then stops automatically. When the user unexpectedly stops speaking, the speech recognizer class automatically stops recording the speech abruptly. Therefore, future plans involve programming the speech recording, where the user controls the starting and stopping of recording, and sending the command as an audio file to the server for speech recognition. Also, the current speech recognition has shown issues such as recognizing words with similar phonetics, combining different spoken words in command, and recognizing as an entirely different phrase. Advanced deep learning-based speech recognition methods will be integrated with the framework for better recognition of user speech commands or user intent. With respect to IoT, the system has to be tested with multiple smartphones.

## REFERENCES

[1] Tzu-Kuan Chuang, Ni-Ching Lin, Jih-Shi Chen, Chen-Hao Hung, Yi-Wei Huang, Chunchih Teng, Haikun Huang, Lap-Fai Yu, Laura Giarré, and Hsueh-Cheng Wang. 2018. Deep trail-following robotic guide dog in pedestrian environments for people who are blind and visually impaired-learning from virtual and real worlds. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5849–5855.

[2] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M Kitani, and Chieko Asakawa. 2019. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *The 21st International ACM SIGACCESS conference on computers and accessibility*. 68–82.

[3] Mohit Jain, Nirmalendu Diwakar, and Manohar Swaminathan. 2021. Smartphone usage by expert blind users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[4] Bing Li, Juan Pablo Munoz, Xuejian Rong, Qingtian Chen, Jizhong Xiao, Yingli Tian, Aries Arditi, and Mohammed Yousuf. 2018. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing* 18, 3 (2018), 702–714.

[5] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2018. Smartphone-based indoor localization for blind navigation across building complexes. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–10.

[6] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2019. Smartphone-based localization for blind navigation in building-scale indoor environments. *Pervasive and Mobile Computing* 57 (2019), 14–32.

[7] Harish Ram Nambiappan, Stephanie Arevalo Arboleda, Cody Lee Lundberg, Maria Kyrarini, Fillia Makedon, and Nicholas Gans. 2022. MINA: A Robotic Assistant for Hospital Fetching Tasks. *Technologies* 10, 2 (2022), 41.

[8] Samuel Poirier, François Routhier, and Alexandre Campeau-Lecours. 2019. Voice control interface prototype for assistive robots for people living with upper limb disabilities. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 46–52.

[9] Lei Wu, Redwan Alqasemi, and Rajiv Dubey. 2020. Development of smartphone-based human-robot interfaces for individuals with disabilities. *IEEE Robotics and Automation Letters* 5, 4 (2020), 5835–5841.

[10] He Zhang and Cang Ye. 2020. A visual positioning system for indoor blind navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9079–9085.