Development and Evaluation of Voice Recognition Input Technology in Navigation System for Blind Person

Anuar bin Mohamed Kassim¹, Takashi Yasuno¹, Hazriq Izzuan Jaafar² and Mohd Shahrieel Mohd Aras²

¹Graduate School of Advanced Technology and Science, Tokushima University
2-1 Minami-Josanjima, Tokushima, 770-8506, Japan
Phone/FAX: +81-88-656-7458
E-mail: {anuar, yasuno}@ee.tokushima-u.ac.jp

²Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Phone/FAX: +60-6555-2358
E-mail: {hazriq, shahrieel}@utem.edu.my

Abstract

This paper describes a navigation system for a blind person that uses an audio guidance system including input from a voice recognition system. The proposed voice recognition system is used in order to easily set the desired destination for the blind person. By using the proposed voice recognition system, the destination can be set by using a microphone. An EasyVR module is used to develop the voice input navigation system. In order to measure the performance of the developed system, the user’s voice is evaluated by using the template matching method. Some experiments on the proposed voice input navigation system are also conducted at different locations with different sound intensity levels. As a result, the satisfactory performance of the developed voice navigation system is confirmed.

1. Introduction

In early 2014, the United Nations (UN) released statistics showing that 10% of the world’s population have a disability, with 80% of them located in developing countries. Moreover, a disability that may involve additional danger to the individual is blindness. Conventionally, blind people rely on a white cane or guide dog to assist them to reach the desired destination [1]. However, these conventional methods are only beneficial if the path to the destination is already known. Navigation becomes difficult if the destination is unknown or the area is not designed for blind people. Since it is unclear where to go and the surroundings cannot be recognized without appropriate navigation tools.

Therefore, advanced technology has been developed by researchers to help blind people to travel independently [2] - [4]. In this paper, a navigation system for a blind person that uses an audio guidance system including voice input technology and a voice recognition system is described. The proposed voice input technology and voice recognition system are applied to enable a blind person to set the desired destination easily compared with input via a Braille keypad. Here, one of the simplest and most accurate methods of voice recognition, the template matching method, is applied in the blind navigation system.

2. Developed Navigation System

Figure 1 shows an illustration of the developed navigation device. In the system, there are some components such as an RFID reader, a microcontroller, a voice module, and an EasyVR module. An Arduino Pro Mini microcontroller is used as an integrated circuit, which contains the memory and the program used to communicate with the other hardware. The RFID detection module (MiFare high frequency, 13.56 MHz), which includes of the RFID reader, is installed at the bottom of a retractable cane for the easy detection of RFID tags installed on tactile paving. Each RFID tag contains prestored information such as the location and the surrounding environment including obstacles, place names, and building names in the library of the microcontroller.

Figure 1: Developed navigation device

The RFID reader reads the tag ID and the code is encrypted and converted into voice commands. Then, the user receives commands on how to travel and information on the surrounding environment. The voice module is used to play the voice commands and inform the user which direction needs to be followed. The voice commands that are played...
are preprogrammed inside the voice module. Previously, a Braille keypad was used in order to set the desired destination. However, it was difficult to set the destination using the Braille keypad, thus it has been replaced by a microphone and the proposed voice recognition system. At the bottom of the cane, a wheel is mounted so that the developed RFID reader can easily detect RFID tags and the user does not need to raise and swing the cane while traveling.


3.1 System construction

Figure 2 illustrates the configuration of the developed navigation system. A microphone acts as the input device in the voice recognition system. A received voice sample is sent to the EasyVR module in order to recognize the voice sample and match it with a prerecorded voice command. The program for the voice recognition system was designed using the Arduino programming language. If the voice command of the user matches a prerecorded voice command, EasyVR will permit the electronic cane to guide the blind person to the destination that has been set. Through the detection of the RFID tag containing information stored electronically, a recorded output voice command will lead the blind person by giving exact directions to reach the destination.

Figure 3 shows a flowchart of the proposed navigation system. Firstly, the RFID reader is initialized. The RFID reader determines the starting point by touching the nearest RFID tag. If the RFID reader fails to read the tag even if it is already in the detection range, the RFID reader is initialized again. In contrast, if the RFID tag is successfully detected, the Arduino microcontroller encrypts the tag identity. Then, the user can determine the desired destination by inputting the location through the microphone. All the destinations have been prestored in the microcontroller library. After the voice is detected by the template matching method, the path localization process leads the user to the desired destination via the voice guidance system, which gives instructions through the single headphone.

3.2 Voice recognition system

Commonly used voice recognition methods are template matching and feature analysis. In this study, a simple location or destination is spoken by the user through a microphone, which is transformed into electrical signals by using an analog to digital converter. Hence, the template matching method was chosen since the template matching method is the simplest technique and has the highest accuracy. However, since the voice of each person varies, the program must be trained with a new user’s voice input, as shown in Fig. 4(a). In the training phase, the user needs to speak a word several times to the microphone and the system extracts the features of the voice and stores a template in the memory using the feature extraction method. Furthermore, during the detection phase, features matching is applied by comparing the database of samples and the input voice, as illustrated in Fig. 4(b). Both feature extraction and feature matching modules are equipped inside the EasyVR module [5].

On the other hand, the amplitude is the property focused on when analyzing the results obtained from an oscilloscope. This is because the higher the amplitude of the voltage, the higher the volume of the sound. Therefore, a difference in amplitude in a sound waveform is caused by the input of sounds with different loudness. The amplitude or peak value which is shown in Fig. 5 is indicated as a voltage. To find
the average voltage $V_{ave}$ of a non-sinusoidal waveform, the area underneath the waveform needs to be calculated. Equation (1) shows the calculation method for the average voltage $V_{ave}$ using the mid-ordinate rule, where $n$ is the numbers of mid-ordinates used and $V_i$ is the peak voltage for each sample time.

$$V_{ave} = \frac{1}{n} \sum_{i=1}^{n} V_i$$  \hspace{1cm} (1)

Figure 5: Average value of non-sinusoidal waveform

4. Experimental Setup

To evaluate the performance of the proposed method, we conducted three experiments. The first experiment was conducted to detect the voice sample for the word "Toilet" and compare it with eight different words: 1: Platform, 2: Station, 3: School, 4: Library, 5: Toilet, 6: Shop, 7: Cafeteria, and 8: ATM. The purpose of this experiment was to compare the waveforms of the eight samples and determine which waveform was detected by the proposed method. The experimental data were obtained by calculating the average voltage of the sample waveform, obtained by an oscilloscope at each sampling time of 40 ms using Eq. (1). The second experiment was carried out using the same method but for the voice sample of "ATM". Both experiments were conducted in a laboratory with a sound intensity level at 50-59 dB, as shown in Fig. 6(a).

The third experiment was conducted in order to determine the accuracy of the proposed method at locations with different sound intensity levels: the laboratory and cafeteria illustrated in Figs. 6(a) and 6(b), respectively. This is because the developed system is expected to be used under different conditions such as noisy places. In the third experiment, only the word "Toilet" is used, which is tested by 30 persons at each location to compare the effects of different sound intensity levels of 50-59 dB (quiet place) and 60-69 dB (noisy place). The indicator of accuracy is the number of successful detections of the voice command.

Figure 6: Sound intensity level in experimental fields

5. Experimental Results

Figure 7 illustrates the result of the first experiment for the original voice of "Toilet" recorded in the EasyVR module, which is compared with the voice samples of eight different words. The average voltage $V_{ave}$ is calculated using the mid-ordinate rule at each sampling time of 40 ms and plotted in Fig. 7. Table 1 shows the result of the voice sample detected for "Toilet" by the proposed method. From the results, the highest amplitude obtained in the entire graph is 6.8 mV, while the lowest is 1.2 mV, which were acquired from the original voice and the eight samples, respectively. The result for the fifth sample, which is the word "Toilet" was detected because of its similarity with the original graph in the memory in terms of the graph flow, in which the maximum voltage rises and drops rapidly.

Table 1: Voice detection result for toilet

<table>
<thead>
<tr>
<th>Word</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

On the other hand, Fig. 8 shows the average voltage versus time graph for "ATM". The original voice recorded in the EasyVR module was compared with the eight voice samples obtained from the different words. Table 2 shows the result of the voice sample obtained for "ATM" that was detected by the proposed system. The highest point in the graph is
reached by sample 5 and the lowest point is reached by sample 2, with peak voltages of 6.8 mV and 1.2 mV, respectively. The graphs of the original sample and eighth sample, corresponding to the word "ATM" have similarities in their shapes: the voltage rises and drops back to the original voltage after a short period and then rises again to the same amplitude as reached before. Figures 9(a)-9(d) show the raw data obtained by the oscilloscope during the experiments when the original words were "Toilet" and "ATM", together with the sample voices of "Toilet" and "ATM".

Table 2: Voice detection result for ATM

<table>
<thead>
<tr>
<th>Word</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 8: Comparison of peak average voltage of samples for ATM

Figure 9: Comparison of voice waveforms for toilet and ATM

Table 3: Performance at locations with different sound intensity levels

<table>
<thead>
<tr>
<th>Range of sound intensity level (dB)</th>
<th>50 - 59</th>
<th>60 - 69</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of successful detections</td>
<td>30</td>
<td>22</td>
</tr>
<tr>
<td>Number of failed detections</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Total number of data</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

6. Conclusion and Future Work

A navigation device for a blind person was developed and the proposed voice input technology including the voice recognition system was evaluated. The tested voice samples were detected successfully by the proposed voice recognition system. The RFID detection system is beneficial for a blind person since it provides feedback information about the current location and the surrounding obstacles. The satisfactory performance of the voice recognition system at different sound intensity levels was also successfully confirmed. In the future, we plan to implement the shortest-path algorithm in the blind navigation device using RFID technology.

Acknowledgment

This research is part of a collaboration project between Tokushima University and Universiti Teknikal Malaysia Melaka.

References


