Original Article :

Improving Outdoor Sound Distance Navigation for Visually Impaired People Using BLSTM and CRF.

Mohsen Bakouri

Department of Medical Equipment Technology, College of Applied Medical Science, Majmaah University, Majmaah City 11952, Saudi Arabia

Department of Physics, College of Arts, Fezzan University, Traghen 71340, Libya

Received on June 03, 2023, accepted for publication on Oct 30, 2023, dx.doi.org/10.5455/mjhs.2024.02.007

Corresponding author:
Mohsen Bakouri,
Department of Medical Equipment Technology, College of Applied Medical Science, Majmaah University, Majmaah City 11952, Saudi Arabia, m.bakouri@mu.edu.sa

OPEN ACCESS Distributed under Creative Commons CC BY-NC-ND 4.0

Abstract

Background and Aims:
This study aims to develop a prototype that utilizes sound-based systems to localize individuals with visual impairments. In this study, the wireless ultrasound network is proposed and used to facilitate autonomous navigation and maneuvering for individuals with visual impairments.

Methods:
In the proposed method, we designed robust algorithms for ultrasonic systems with sound waves of high frequency to detect obstacles in the surrounding environment. The algorithms were developed using voice recognition method, bi-directional long short-term memory (BLSTM), and conditional random fields (CRFs). The proposed system was implemented with the help of assisting hardware instruments, such as an ultrasonic sensor network and a global positioning system (GPS). To analyze different outdoor settings, the Raspberry pi4 was used to locate and store the coordinates (points of interactive latitude and longitude) of four different outdoor places. These including home, mosque, laundry, and supermarket.

Results:
According to our results, the proposed approach was capable of accurately navigating the specific site with an accuracy of 98.5%.

Conclusions:
The proposed method was able to obtain a robust outdoor localization for VIP.

Keywords:
Visually Impaired People, Outdoor Navigation, Long Short-Term Memory, Bi-directional long short-term memory, Conditional random fields.

Introduction

Various technologies are presently utilized to augment the mobility of individuals who are blind or visually impaired.[1] The sensors provide limited data regarding individuals moving between different places.[2,3] The array of outdoor sensing devices includes a range of technologies such as cameras, intelligent navigation, and GPS tools. These devices have the capability to input a more significant amount of information compared to indoor methods. This data may relate to the surroundings covering individuals with visual impairment. In addition, it should be noted that the cost of these devices is higher compared to their indoor counterparts owing to the inclusion of a processing phase that necessitates the use of computers or microprocessors.[4, 5] visually impaired people (VIP) commonly encounter obstacles while traversing their environment, encompassing interior and exterior spaces. Implementing localization strategies for individuals with visual impairments facilitates their ability to navigate, ensures their well-being, and promotes self-sufficiency. Blind individuals may not always have access to conventional localization methods, such as GPS or visual cues. To address these issues, several investigations have been considered.[6–8] For instance, Nagarajan et al. proposed a method that utilized low-power Bluetooth-enabled devices to ascertain the precise coordinates.[9] The study used an algorithm with different data collected and conducted an experimental analysis to demonstrate the most appropriate placement of maneuvers. In addition, the algorithm was executed in the form of an Android application, which offered a navigation system for VIP independent of any external resources.

Localization techniques that are conducted based on machine learning usually utilize information from several sources, including sensors, to gather information about the behavior and preferences of the user.[10] The system can employ the data to offer the user individualized recommendations and guidance.[6, 11] For example, Ashiq et al. developed a convolutional neural network (CNN) model to direct VIP.[12] The model utilized an online tool for object identification and recognition methods. By using this feature inside the application, the user was allowed to share their location with family and then was able to track down the VIP and determine the exact location. A different investigation was carried out by Tan et al. using an interaural phase difference localization method to detect the direction and distance of the person.[13] The method was testing with people who are blind or have visual impairments. The results demonstrated that the system was

effective in conveying spatial information.

Despite the numerous benefits of the above methods, the challenge of achieving real-time sensing still needs to be solved and developed. Navigating through unfamiliar locations poses a significant challenge for individuals with VIP. In addition, VIP may not always have access to standard localization methods like global positioning system (GPS). For this reason, different researchers have been looking into various strategies that provide spatial information using many senses.[14] Modern technological advancements, exemplified by the incorporation of sensors in smartphones, have the capability to recognize intricate surroundings, explore novel locations, and provide auditory guidance to users for navigation purposes.[15,16] However, there is a need to improve the efficacy of these strategies, particularly in circumstances that involve complex environments. Robots have the potential to successfully complete assignments in terms of their capabilities and can achieve all the necessary goals. Automated approaches have been shown to be highly effective in the completion of challenging navigation tasks, when applied to unfamiliar environments in both global and local contexts. Despite the fact that robots can provide useful information about impediments, their application is still limited in regional and global markets, and they are currently undergoing testing in clinical settings.[17,18]

The development of modern technologies for VIP is an important necessity for researchers to explore the availability of accurate and reliable spatial data on an ongoing basis. This study aimed to develop and execute a sound-based localization methodology that could effectively navigate and orient visually impaired individuals to their intended destination in a timely manner. This methodology involves the utilization of spatialized sound for the user with pertinent details regarding their current location and environment. The individual utilizes headphones or earbuds, while the system delivers auditory directives contingent upon their spatial positioning. Safety and real-time processing factors were taken into account during design to attain autonomous mobility, detect impediments encountered by visually impaired individuals in indoor settings, and adapt to novel, intricate surroundings.

Materials, and Methods

The system architecture

This study presents an efficient and manageable electronic guidance system, facilitating the navigation of visually impaired individuals within pre-established locations. The method was integrated with a GPS, to guide individuals toward pre-de-

termined destinations. This study involved using voice commands with hot keywords by VIP in an outdoor scenario. The outdoor system was equipped with voice commands that includes the system's orientation as well as the four different locations' coordinates (the supermarket, the mosque, the laundry, and the house). A GPS, a digital compass, a speaker, and a microphone were all included in the range of outdoor equipment. The proposed system's prototype is depicted in Figure 1. The components specifications of hardware for the proposed system as given in Table A.1.

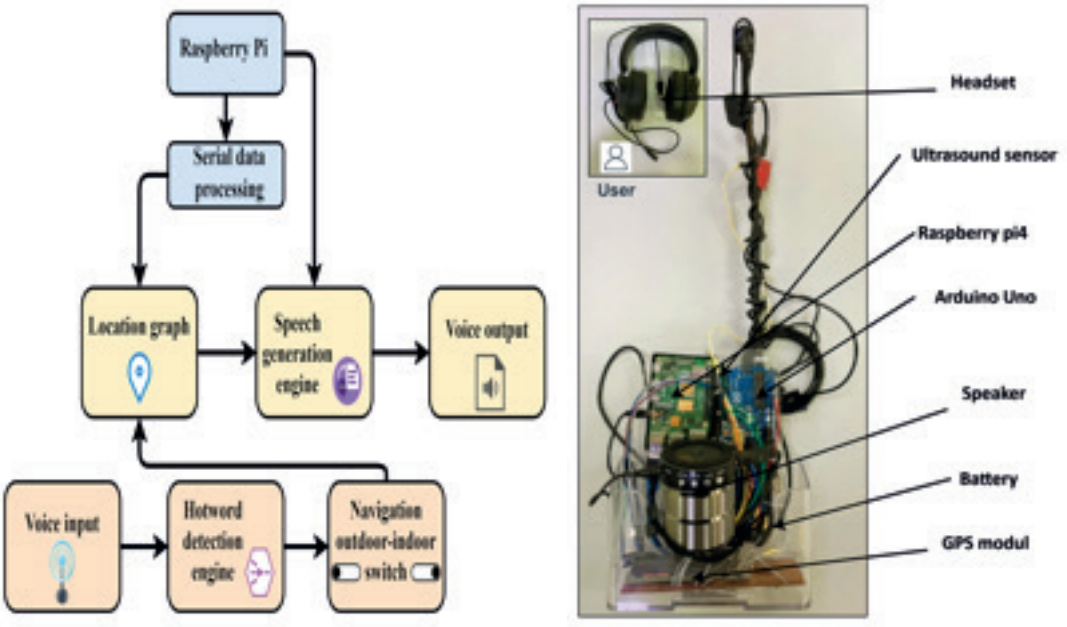


Fig. 1. Structure of proposed model.

Outdoor navigation algorithm

In the current study, a VIP uses voice commands to ask for a specific set of position coordinates by speaking into a microphone with a predetermined keyword. The audio file is then delivered to the system, where it is processed using voice recognition software. The voice file is transformed into a text file. The software program will conduct a comparative analysis between the textual documents and the designated name of the location that was previously stored. By, assuming the software successfully identifies a corresponding value, it gathers the geographical coordinates. Subsequently, it transmits this data to the visually impaired individual via auditory recordings, providing navigational instructions for reaching the destination. GPS is utilized for real-time global positioning determination. The digital compass is utilized to ascertain the present orientation in actual time. The system assists the visually impaired individual in navigating to their intended destination by providing directional information through a wireless

headset using auditory tools. Furthermore, the visually impaired individual is notified regarding the closest objects.

In this method, voice recognition methodology generated a frequency distribution for every audio file. The utilization of the bi-directional long short-term memory (BLSTM) model in combination with conditional random fields (CRFs) was also employed for the identification and filtration of the output files as described below.

Voice Recognition Approach

Figure 2 illustrated the voice recognition approach that used to implement this method. In this approach, Mel-frequency cepstral coefficients (MFCC) were utilized by the present methodology to derive the feature map of the data included within the audio file.19 The extraction method included data normalization and the application of a finite impulse response filter (FIR) to each audio file, which can be mathematically represented by the following equation:

$$y(n)=v(n)-\beta v(n-1) \quad (1)$$

where v(n) is the filter output, v(n) is the audio file, n is the number of samplings, and β is given as $(0.9<\delta\leq 1)$.

Framing and windowing techniques were utilized in order to mitigate signal discontinuity as:

$$\omega(n) = \begin{cases} 0.5 - 0.45(1 + \cos(2\pi n/l - 1)) & 0 \leq n \leq l-1 \\ 0 & otherwise \end{cases} \quad (2)$$

where l is the number of frames.

The spectrum magnitude of each frame was determined by applying the Fast Fourier Transform (FFT) with complexity (N*log(N)), as expressed in the following equation:

$$V(k)=\sum_{n=0}^{N-1} v(n)e^{-j2\pi kn/N}, \quad (3)$$

Subsequent to the FFT procedure, the outcomes were subjected to the Mel Frequency (Y_t (k)) analysis using Mel-scale algorithm as:

$$Y_t(k) = 2595\log_{10}((700+f)/700) \quad (4)$$

The following equation can be used to compute the logarithmic of the square magnitude of the Mel-filter bank's output as:

$$y_t'(k) = \log |Y_t(k) * V(k)|^2 \quad (5)$$

In this study, we inverse DFT to calculate Mel Frequency spectrum as:

$$y_t(j) = \sum_{k=1}^{M} \log(|Y_t(k)|)\cos\left(\emptyset(k-0.5)\frac{\pi}{M}\right), \emptyset = 0,1,..,j \quad (6)$$

As speech signal is not constant, the dynamic cepstral coefficient is added as energy feature:

$$e_t = \sum_{t=t_1}^{t_2} v^2(n) \quad (7)$$

Finally, first and second derivative are applied in order to in order to obtain temporal information as:

$$y_t = \begin{Bmatrix} y_t(j), e_t \\ \Delta\{y_t(j)\}, \Delta\{e_t\} \\ \Delta^2\{y_t(j)\}, \Delta^2\{e_t\} \end{Bmatrix} \quad (8)$$

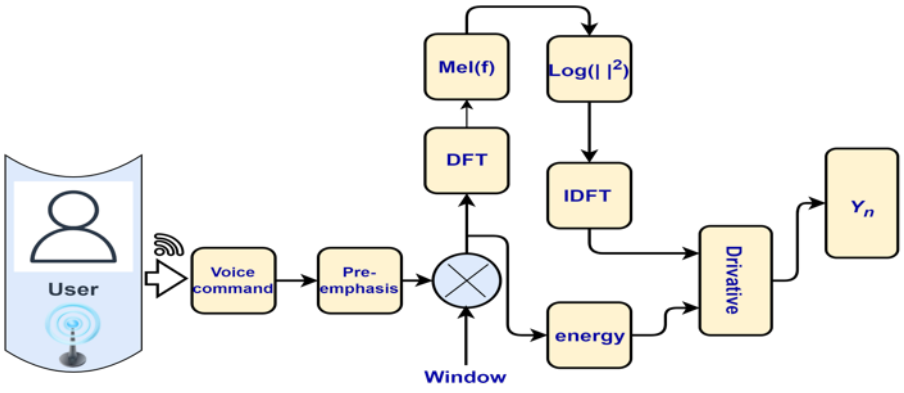


Fig. 2. Voice recognition approach.

BLSTM Classification Adoption

Bi-directional long short-term memory (BLSTM) layers consist of a pair of distinct LSTM networks as shown in Figure 3. One of these networks processes the input in a forward direction, moving from the beginning to the end. The other network operates in a backward direction, moving from the end to the beginning. The ultimate output is generated by concatenating the concealed states of the two LSTMs that are oriented in opposite directions. [20]

The LSTM unit is endowed with several gates that facilitate the regulation of information flow. The fundamental components of an LSTM cell are the input, output, and forget gates. The input gate denoted by (i_t) and responsible for overseeing the incoming data and given by:

$$i_t = \sigma(W_{h_i} h_{t-1} + U_{x_i} x_t + b_i) \quad (9)$$

while the forget gate (f^t) is tasked with determining which aspects of the previous memory dynamics should be discarded and given by:

$$f_t = \sigma(W_{h_f} h_{t-1} + U_{x_f} x_t + b_f) \quad (10)$$

the output gate, denoted by (o_t), plays a crucial role in determining which specific piece of information should be allowed to flow out of the cell and calculated as:

$$o_t = \sigma(W_{h_o} h_{t-1} + U_{x_o} x_t + b_o) \quad (11)$$

the cell state value (c_t) is updating throughout the intermediate value of the cell as:

$$\tilde{c}_t = tanh(W_{h_c} h_{t-1} + U_{x_c} x_t + b_c) \quad (12)$$

where, x_t is the input vector, h_t is the state hidden, W_h is the weight matrix for the hidden state, U_x is the weight matrix for the gates, b is the biases for the hidden state.

here the c_t h_t and are given by

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

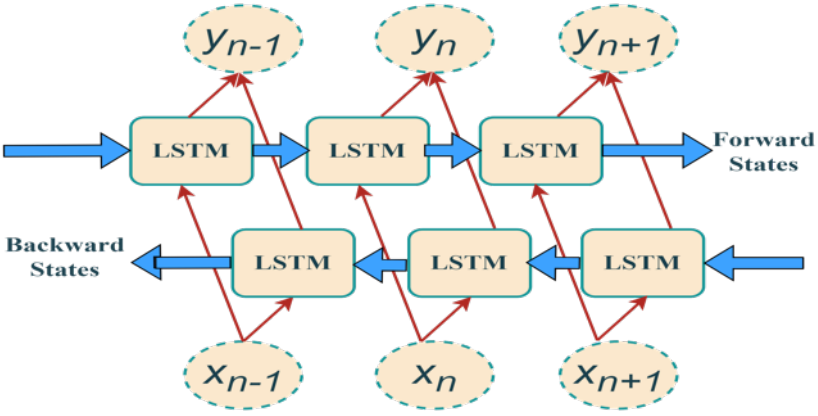


Fig. 3. BLSTM of proposed model

CRF Classifier Adoption

Conditional random fields (CRFs) are undirected graphical models and are utilized to calculate the probability of output nodes, given the values present in the input nodes.[21] The individuals can collaboratively anticipate the optimal classification for every constituent within the given sequence of inputs. Therefore, to solve the task of sequence labeling, it is necessary to consider the interdependence among adjacent elements to accurately assign labels to individual elements within a provided input sequence. To predict the best label for each element in the input sequence, we proposed the following CRF model as:

$$\rho_\delta(y|x) = \frac{1}{Z(x)} \prod_\gamma \mu_\gamma(y_\gamma, x; w) \quad (15)$$

where, y is the output of the CRF model sequence, x is the observed variable, y_γ is the predicted variable, w is represented the set of parameters, and $\mu_\gamma (y_\gamma,x;w)$ is the potential function and given by:

$$\mu_\gamma(y_\gamma, x; w) = \exp(w.f_\gamma(y_\gamma, x)) \quad (16)$$

the term $Z(x)$ known as normalization function and given by:

$$Z(x) = \prod_y exp(w.f_\gamma(y_\gamma, x) \quad (17)$$

By substituting eq (16) in (15) we can write:

$$\rho_\delta(y|x) = \frac{\exp(w.f_\gamma(y_\gamma,x))}{\prod_y exp(w.f_\gamma(y_\gamma,x)} \quad (18)$$

One possible approach to setting objectives involves assigning a high probability to observed outcomes and a low probability to all other potential outcomes. For this purpose, the logarithm of the likelihood $L(w)$ is utilized to estimate the parameters in the following manner:

$$L(w) = \sum_{y \in D} log\rho_\delta(y|x) \quad (19)$$

The computation of the gradient of the objective function is performed as:

$$\frac{dL(w)}{dw} = \sum_i f_\gamma(y_\gamma, x) - \sum_i \rho_\delta(y|x) \sum_y f_\gamma(y_\gamma, x) \quad (20)$$

The estimates of optimal label sequence (y^*) can be derived from the maximum-likelihood method are:

$$y^* = \underset{y \in R^D}{\operatorname{argmax}} \sum_{y \in D} log \rho_\delta(y|x) \quad (21)$$

The structure of the CFR used in the proposed system was illustrated in Figure 4.

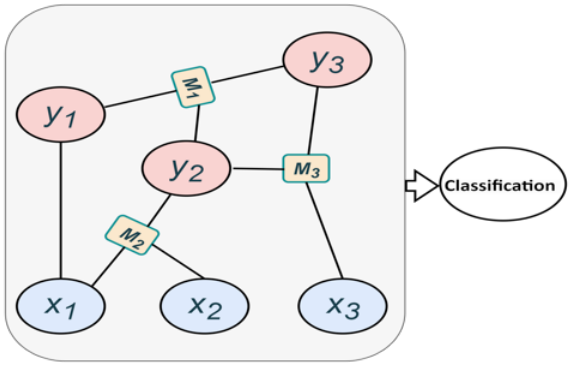


Fig. 4. Structure of CFR model

Evaluation method

Algorithms were designed and evaluated using Python and C++ software.22 The proposed method was evaluated using an English speech group comprising distinct words. The precise enunciation of entire set of 5000 words ascertained through the participation of twelve Arabic fluent speakers. The data was sampled at a frequency of 25 kilohertz and had a bit resolution of 16. The parameters of speed, dynamic range, noise, and temporal shifts in both forward and backward directions were subsequently modified. Around 80% of the specimens, totaling 4000, were allocated to the development of the training set, which included both training and validation phases. The remaining 20% of the samples, amounting to 1000, were designated for the creation of the test set. The experimental procedures were conducted over a span of 50 epochs, with each batch consisting of four participants. To evaluate the proposed system's predictive quality and accuracy, we calculated the F-score by incorporating precision (p) and recall (r) through the following formula

$F = 2[p * r/(p + r)]$, where $p = [t_p/(t_p + f_p)]$ and $r = [t_p/(t_p + f_n)]$; here, t_p, f_p, and f_n are the true positive, false positive, and false negative, respectively. The real experiments was take place in Al-Arid district, Riyadh City.

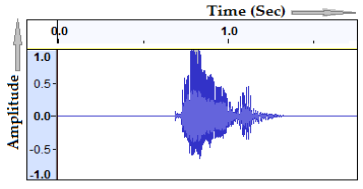
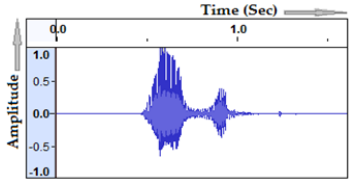
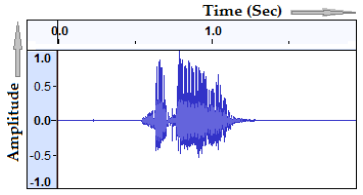
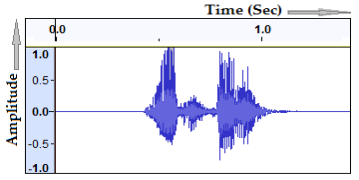
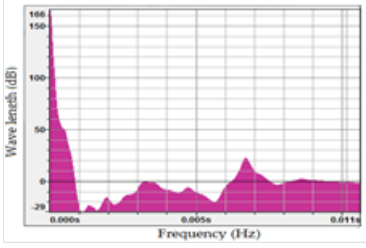
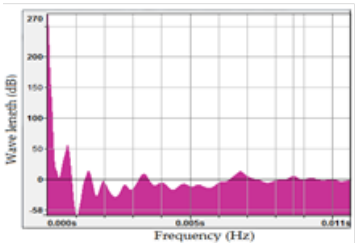
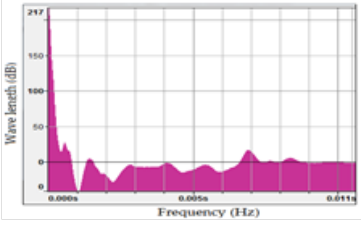
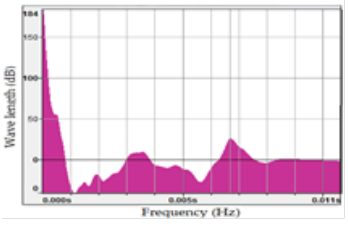
Results

MFCC step has generated the waves as shown in Table 1.a shows the four audio wave shapes with training time. It ca be seen that the training times are 5s for all waves. As indicated in Table 1.b, the audio file waves need to be transformed into their frequency domain using Fourier analysis. Then the system used convert the frequency domain waves to time-frequency spectrogram as illustrated in Table 1.c. Time-frequency spectrogram waves

were used as input to the BLSTM and CRF models respectively for final classifications to produce the confusion matrix shown in Table 2.

From Table 2. the average accuracy all five voice commands was approximately 98.5%. The computations results of accuracy, precision, recall, and F-Score are disapplied in Table 3.

Tab. 1. MFCC step of generated the waves.

<table>
<tr><td rowspan="4">a. audio wave shapes.</td><td>Mosque</td><td>Training time=0.5s</td><td>Laundry</td><td>Training time=0.5s</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td>Supermarket</td><td>Training time=0.5s</td><td>Home</td><td>Training time=0.5s</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td rowspan="4">b. singles of long-term spectrum.</td><td>Mosque</td><td>Training time=0.5s</td><td>Laundry</td><td>Training time=0.5s</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td>Supermarket</td><td>Training time=0.5s</td><td>Home</td><td>Training time=0.5s</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
</table>

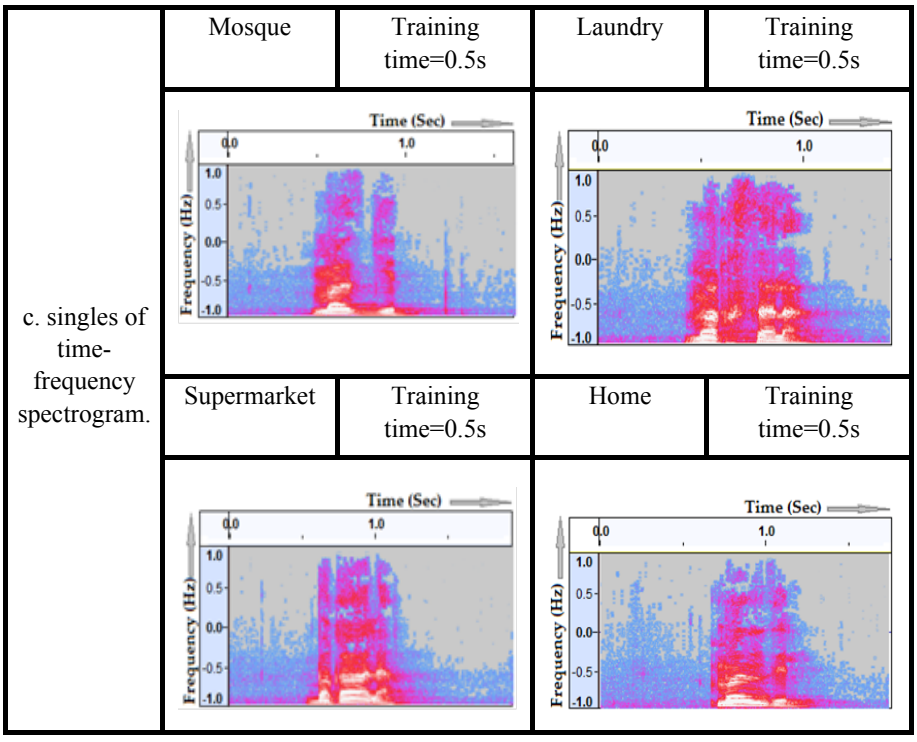


Table 2: Confusion matrix.

Voice Command					
Prediction ratio %	Class	Mosque	Laundry	Supermarket	Home
	Mosque	98%	0.5%	0.9%	0.8%
	Laundry	0.5%	98.5%	0.9%	0.9%
	Supermarket	0.9%	0.9%	99%	0.7%
	Home	0.8%	0.9%	0.7%	98%

Table 3: The evaluation metrics of accuracy, precision, recall, and F-score.

Class	Accuracy	Precision	Recall	F-Score
Mosque	98%	0.85	0.85	0.90
Laundry	99.3%	0.90	0.86	0.80
Supermarket	98%	0.87	0.87	0.85
Home	99%	0.87	0.89	0.87

Fig. 5 Shows navigation in outdoor map (home to supermarket) at (Al-Arid district, Riyadh City). The individual with visual impairment attempted to navigate from their residence to the supermarket area, as depicted in Figure 5, utilizing the coordinates specified in Table 4. The system generated a graphical representation comprising 26 interconnected points, denoting nodes, from the initial node, designated as "home," to the terminal node, labeled as " supermarket." The inter-node distance ranged from 5 to 10 meters. The visually impaired individual was provided with guidance on progressing along the route. The system provided the blind individual

with location information and guidance while accompanying family members followed the individual throughout the journey. The individual with visual impairment ambulated while receiving GPS data and digital compass navigation through a headset. A record was kept of all the nodes traversed by individuals with visual impairments.



Fig. 5. Navigation in outdoor map (home to supermarket).

Table 4: Example for destination coordinates.

Planned longitude	Planned latitude	Actual longitude	Actual latitude
25.894784	47.61832	25.894799	47.618373
25.894765	47.61833	25.894795	47.618381
25.894748	47.61834	25.894792	47.618389
25.894731	47.61835	25.894787	47.618387
25.894712	47.61836	25.894712	47.61836
25.894673	47.61838	25.894699	47.618398
25.894612	47.61847	25.894662	47.618454
25.894595	47.61841	25.894595	47.618442
25.894522	47.61846	25.894572	47.618477
25.894503	47.61847	25.894503	47.61847
25.894484	47.61848	25.894494	47.618493
25.894406	47.61852	25.894456	47.618555
25.894326	47.61846	25.894366	47.618499
25.894311	47.61842	25.894361	47.618446
25.894296	47.61839	25.894296	47.618399
25.894272	47.61835	25.894262	47.618389
25.894231	47.61831	25.894271	47.618339
25.894216	47.61826	25.894266	47.618291
25.894148	47.61808	25.894178	47.618081
25.894129	47.61803	25.894129	47.61803
25.89408	47.61803	25.89408	47.618025
25.894034	47.61804	25.894094	47.618061

25.893908	47.6181	25.893948	47.618142
25.893872	47.61812	25.893892	47.618151
25.893831	47.61815	25.893891	47.618178
25.893804	47.61817	25.893854	47.618189

Discussion

This research was conducted with the creating a reliable and cost-effective system to guide and assist those who are visually impaired. The primary motivation behind this design was to give blind individuals a far higher degree of autonomy and, by extension, a better quality of life. To improve upon the current system design, the proposed smart wearable prototype incorporates state-of-the-art smart control systems such voice recognition, BLSTM model, CRF model, and GPS navigation. Voice recognition systems, wireless network infrastructures, and major developments in sensor technologies are all responsible for the widespread usage of navigation technology for directing blinds. The suggested system is distinguished by its low cost, low energy consumption, and the simplicity of installing the electrical and electronic circuits. An example of a simple electronic circuit connection is depicted in Figure 1. The materials and methods employed, as well as their adaptability, personalization, and dissemination to the end user, make the design both efficient and affordable. If you process one assignment at a time, you can go around every obstacle in roughly half a second. This intelligent prototype's software can all function without an internet connection. The proposed software is also highly accurate, even when operating in an environment with background noise.

Background noise from numerous sources may have a negative impact on sound-based localization systems for the blind. As a result, the system's capacity to pinpoint the origin of the sought-after sound can be improved even in noisy environments. The current research targeted the development of voice recognition algorithm techniques employing MFCC, BLSTM, and CRF models to strengthen the proposed system. When the signal-to-noise ratio (SNR) was at least 14.8 dB, the investigation showed that the system attained accuracy of 98.5%. In contrast, when contrasting several localization strategies, it is common practice to utilize a quantitative criterion, such as accuracy, to evaluate the relative merits of each strategy. The efficiency of many methods can be compared using this metric so that the best one can be chosen for a given task.

Based on comparisons to prior research on efficacy, reliability, and cost, we believe that the design and implementation tech-

nique we used in this study has generally addressed a number of issues. For instance, a recent work by our research team tested and used an assistive wearable smart system for VIPs that uses ultrasonic sensors to navigate both indoor and outdoor settings while avoiding impediments.[19] The design used the Dijkstra algorithm to calculate the shortest distance between two locations in order to carry out the performance with high precision. Additionally, this system uses LSTM and MFCC to attain accuracy of 97%. This study also determines the power consumption of a system made for outdoor localization to determine how long it will function when included into an assistive device. As part of analyzing the system's power usage, measurements are taken for outdoor conditions. The results showed that the inside localization used 1.3 watts on average, whereas the outdoor localization used 1.8 watts on average. Furthermore, the analysis shows that the system can operate for about 11 hours outside, based on the assistive device's battery capacity, which is utilized to compute the system lifetime.

Although the results of this study show that the assistive devices can be created a robust localization system that work effectively over time in various settings. However, this study needs more actual tests by experimenting different sounds and also testing in the closed spaces.

Conclusions

The present study involved the development of a hybrid intelligent wearable navigation prototype that can facilitate automatic guidance for individuals with visual impairments in outdoor environments. The proposed prototype was implemented through the utilization of software and hardware tools. This method was implemented through various assistive hardware tools, such as the Raspberry Pi 4, ultrasound sensors, the Arduino Uno microcontroller, a GPS module, and a headset. The robust techniques used to develop the Python and C++ software to control hardware elements via an offline Wi-Fi hotspot were created. In addition, MCFF, BLSTM and CRF models were developed to construct an integrated voice recognition system capable of training and distinguishing multiple voice commands, including but not limited to the mosque, laundry, supermarket, and home.

In the study, simulation procedures and evaluation techniques were used to conduct an analysis of a corpus consisting of 5,000 different word utterances that were sourced from twelve Arabic speakers who were fluent in the language. The information was sampled at a rate of 25 kHz and had a resolution of 16 bits when it was

recorded. The accuracy, precision, recall, and F-score of each voice command were calculated using a normalized confusion matrix. According to the findings of the empirical research, the algorithms used in controlled outdoor navigation display a significant level of accuracy with a score of 98.5%. In conclusion, the prototype that has been developed is characterized by its simplicity, affordability, autonomy, and security, among other advantageous features.

Conflict of Interest

None

Funding

None

References

1. Paiva S, Gupta N. Technologies and systems to improve mobility of visually impaired people: A state of the art. Technological Trends in Improved Mobility of the Visually Impaired. 2020:105-23.
2. Tapu R, Mocanu B, Zaharia T. Wearable assistive devices for visually impaired: A state of the art survey. Pattern Recognition Letters. 2020 Sep 1;137:37-52.
3. Plikynas D, Žvironas A, Budrionis A, Gudauskis M. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. Sensors. 2020 Jan 23;20(3):636.
4. El-Taher FE, Taha A, Courtney J, Mckeever S. A systematic review of urban navigation systems for visually impaired people. Sensors. 2021 Apr 29;21(9):3103.
5. Romlay MR, Toha SF, Ibrahim AM, Venkat I. Methodologies and evaluation of electronic travel aids for the visually impaired people: a review. Bulletin of Electrical Engineering and Informatics. 2021 Jun 1;10(3):1747-58.
6. Grumiaux PA, Kitić S, Girin L, Guérin A. A survey of sound source localization with deep learning methods. The Journal of the Acoustical Society of America. 2022 Jul 1;152(1):107-51.
7. Yassin A, Nasser Y, Awad M, Al-Dubai A, Liu R, Yuen C, Raulefs R, Aboutanios E. Recent advances in indoor localization: A survey on theoretical approaches and applications. IEEE Communications Surveys & Tutorials. 2016 Nov 29;19(2):1327-46.
8. Zhang D, Han J, Cheng G, Yang MH. Weakly supervised object localization and detection: A survey. IEEE transactions on pattern analysis and machine intelligence. 2021 Apr 20;44(9):5866-85.
9. Nagarajan B, Shanmugam V, Anantha-

narayanan V, Bagavathi Sivakumar P. Localization and indoor navigation for visually impaired using bluetooth low energy. InSmart Systems and IoT: Innovations in Computing: Proceeding of SSIC 2020 (pp. 249-259). Springer Singapore.

10. Tang Y, Chen M, Wang C, Luo L, Li J, Lian G, Zou X. Recognition and localization methods for vision-based fruit picking robots: A review. Frontiers in Plant Science. 2020 May 19;11:510.
11. Desai D, Mehendale N. A review on sound source localization systems. Archives of Computational Methods in Engineering. 2022 Nov;29(7):4631-42.
12. Ashiq F, Asif M, Ahmad MB, Zafar S, Masood K, Mahmood T, Mahmood MT, Lee IH. CNN-based object recognition and tracking system to assist visually impaired people. IEEE Access. 2022 Jan 31;10:14819-34.
13. Tan TH, Lin YT, Chang YL, Alkhaleefah M. Sound source localization using a convolutional neural network and regression model. Sensors. 2021 Dec 1;21(23):8031.
14. Al-kafaji RD, Gharghan SK, Mahdi SQ. Localization techniques for blind people in outdoor/indoor environments. InIOP Conference Series: Materials Science and Engineering 2020 Mar 1 (Vol. 745, No. 1, p. 012103). IOP Publishing.
15. Mai C, Xie D, Zeng L, Li Z, Li Z, Qiao Z, Qu Y, Liu G, Li L. Laser Sensing and Vision Sensing Smart Blind Cane: A Review. Sensors. 2023 Jan 12;23(2):869.
16. Messaoudi MD, Menelas BA, Mcheick H. Review of Navigation Assistive Tools and Technologies for the Visually Impaired. Sensors. 2022 Oct 17;22(20):7888.
17. Khan S, Nazir S, Khan HU. Analysis of navigation assistants for blind and visually impaired people: A systematic review. IEEE access. 2021 Jan 18;9:26712-34.
18. Simões WC, Machado GS, Sales AM, de Lucena MM, Jazdi N, de Lucena Jr VF. A review of technologies and techniques for indoor navigation systems for the visually impaired. Sensors. 2020 Jul 15;20(14):3935.
19. Bakouri M, Alsehaimi M, Ismail HF, Alshareef K, Ganoun A, Alqahtani A, Alharbi Y. Steering a robotic wheelchair based on voice recognition system using convolutional neural networks. Electronics. 2022 Jan 5;11(1):168.
20. Chakladar DD, Dey S, Roy PP, Dogra

DP. EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. Biomedical Signal Processing and Control. 2020 Jul 1;60:101989.

21. Wang M, Cheng JC. A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. Computer-Aided Civil and Infrastructure Engineering. 2020 Feb;35(2):162-77.