Voice Assistant AI Enabled Smart Cap for Individuals with Visual and Auditory Impairment

[1]Chandra Sekhar Nandi, [2]Snehashish Banerjee, [2]Sourodeep Mallik, [2]Adrija Choudhury, [2]Cynthia Das Bera, [2]Rakib Ikbal Sarkar

[1]Assistant Professor with the Applied Electronics and Instrumentation Engineering Department, University Institute of Technology, The University of Burdwan, Burdwan, India, e-mail: csnandi@uit.buruniv.ac.in.
[2]UG students with the Applied Electronics and Instrumentation Engineering Department, University Institute of Technology, The University of Burdwan, Burdwan, India.

Abstract : With rising cases of visual impairments all over the world, and the number of mishaps occurring on regular basis because of the same, it becomes really challenging for those who face it, be it with their day-to-day chores or with their rather less-important task. Keeping this in mind, we have proposed an Artificial Intelligence (AI) based voice assistant Smart Cap for the people with visual impairment that can aid in multiple tasks. This proposed device is designed and developed to guide its user along their way by making them aware of any obstacle in the vicinity, the state of the traffic lights and also, through the narration of the text to the user. The device essentially comprises a number of input sensors, a microcontroller for processing input information and an output module to provide verbal instructions to the user based on the input signals, accompanied by a physical vibration output. The prototype of this one-of-a-kind device is practically tested and the outcome is promising. The YOLO v3 module, trained with COCO dataset enables this device to recognize objects effectively. In addition, the basic and user-friendly design of this wearable 'cap' makes it easy for its users to wear as well as operate.

Index Terms - – Visual impairment; Auditory impairment, Smart Cap; Artificial Intelligence; Object recognition; Voice assistant.

I. INTRODUCTION

According to a report of World Health Organization (WHO), globally, at least 2.2billion people have a near or distance vision impairment [1]. The research estimates, that eighty to eighty five percent of our perception, learning, cognition and activities are mediated through vision [2], it is therefore quite apparent how difficult the life of an individual is, who doesn't have a proper visual perception. This problem is not only limited to the elderly, but greatly affects children as well. Between October 2018 and December 2019, authors conducted a study among 1,969 students aged <16 years enrolled in 39 schools for the blind, in Maharashtra state, India and found that, 188 children (9.5%) had severe visual impairment and 1,666 children (84.6%) were totally blind [3].

An obstacle detection and avoidance system proposed [4] to assist the blind as well as visually impaired people leveraging an on-board RGB-D camera and Kalman filter algorithm. The system ensures navigation safety by detecting dynamic obstacles in real time and adjusting the path planning. A wearable navigation device was proposed [5] to provide automatic navigation for the visually impaired person. Their device used a depth camera for obstacle detection. The proposed device can be effective for the partially or totally blind person. But the system has no relevance to the auditory impairment person.

The paper [6] focused on the feasibility of using an electromagnetic sensor to assist visually impaired and blind individuals in autonomous walking. They utilized a microwave radar integrated into a traditional white cane to detect obstacles. The system aims to offer better performance, noise tolerance and reduced dimensions. However, the observation range requires extension up to 5m to cover a broad dimension. Authors [7] proposed an electronic Vibration Alert Bracelet (VAB) unit that alerts user through vibration and lighting. The major limitation of this device is that, it is vulnerable to typical security threats, since the proposed VAB system is connected via Wi-Fi.

An AI-based visual aid system for the completely blind people was proposed using Raspberry Pi 3 Model B+ to carry out the distance measurement between obstacle and user, image-to-text conversion with auditory feedback [8]. The system is compact in size, cost effective and showed promising experimental outcomes. The system, however, lacks the advanced features, such as wet-floor, staircases detection, and utilization of GPS and mobile communication module.

A prototype device proposed [9] that comprises a ring consists of several electronic modules. Computer-vision analysis of the camera image controls vibrotactile feedback, leading users to move their hand to near targets. The major disadvantage of this device is it does not support rapid ballistic reaching, as occurs during visually directed action. The system could further be improved by adapting its features to the individual user.

Authors [10] studied the role of wearable devices in improving social interactions among individuals with visual impairments. They provided valuable insights into the transformative potential of wearable assistive devices through the integration of advanced technologies such as eye tracking and artificial gaze. But, their examinations on test subjects should consider their age, educational background and vernacular for improved questionnaire round outcomes.

A device developed using ultrasonic sensor [11] that combines the features of sound and vibration alert. The sensor can detect obstacles within the stipulated range (150cm) and the vibration can be issued by putting user's finger on a button to indicate a risk. The major limitation of this approach is that, touching the button is necessary in order to feel the vibration. The vibration of the motor could also be irritating for the users. The device is incapable of object recognition as it does not contain any camera module.

Recently, the research paper [12] was carried out with smart glasses which utilize the computer vision to detect objects using MS COCO dataset and trained a Convolutional Neural Network model. The system has remarkable feature of face recognition. Their model of OCR for different languages can be applied to give multi-language advantage to the user in future. However, the system's accuracy falls to 50% while recognizing masked or covered faces.

In the research paper [13], authors proposed an intelligent blind guidance system that is capable of detecting traffic lights, effectively prevent obstacles and speech interaction. The major limitation of this device is that requires a constant Wi-Fi connection to function. Recently, authors developed a wearable target location system [14] to facilitate the spatial cognition of blind and visually impaired people. It consists of a head-mounted RGB-D camera, Text-to-Speech system and stereo earphone that provide both VI based object recognition feedback and SAR (Spatial Audio Rendering) based location assistance. Although the device showed a significant success rate, it still needs improvements to overcome the deviation from the actual orientation of objects.

This paper discusses the design and development of a 'Smart Cap' for the individuals with visual and auditory impairment. This prototype device was practically tested and had shown promising performance accuracy. The design of the proposed Smart Cap, experimental set up and result are discussed in section II and III, respectively. Finally, section IV concludes the article with discussions on the shortcomings of proposed device to provide guidance for future developments.

II. DESIGN OF SMART CAP

Smart cap is a device dedicated to the service of a visually and audibly impaired individuals. The system is designed in such a way that it can guide its user in their day-to-day activities. The aspects of proximity and constant audio enabled guidance have been kept in mind, while making this device.

Smart Cap incorporates advanced AI algorithm, ATmega328P microcontroller (an 8-bit RISC microcontroller, maj-SR04 ultrasonic sensor for distance measurement, ESP-32 CAM connected with FTDI module for object recognition, an 8GB SD card containing network credentials, Grove speaker or component of Arduino UNO) for processing purpose, various sensors and actuators to realize its operation, such as, HC for audio output, a vibration motor for generating physical vibration alerts, an LED indicator, power supply unit and a switch. The 'cap' is utilized to accommodate these components.

The step-wise process flow of the Smart Cap is illustrated as a flow chart in Fig.1. The basic algorithm of the working of the device is quite simple-

Step 1: The sensors (ultrasonic sensor, CAM module) collect relevant information about its surrounding and convert them into electrical signal suitable for transmission.

Step 2: Communication between input devices and the microcontroller (Arduino UNO) is set up via suitable means.

Step 3: The microcontroller continuously processes the signal for safety limits, i.e., the signal from the ultrasonic sensor is analysed for safe distance of approaching objects and user. Furthermore, turning off the vibration motor when the signal is red indicates that cars have stopped, and the user can securely cross or move across the street.

Step 4: Machine learning algorithm is put into action.

Step 5: Based on the processed information, the output devices (speaker, vibration motor) are activated.

Step 6: The entire process is looped over.

A schematic representation of the workflow is shown in Fig.2. The signals from the input devices pass to the output device via the microcontroller, which acts as the sole decision maker. The AI camera relies on COCO dataset to keep track of vehicles, traffic lights and other obstacles that help the user in travelling along the busy street.

To increase the reliability of the cap, an ultrasonic sensor is mounted to keep a constant check on the objects approaching the bearer of the device, and when the safety limit is hampered, the speaker and the motor send an alert via audio message and vibration. The audio messages are pre-recorded and converted into an encoded stream of decimal digits with 'Audacity' software. Depending on the distance and type of obstacle, a particular stream is decoded by the processor to play an audio alert.

Fig.3 shows the detailed circuit connection of the proposed device. The exact connections have been shown upon checking in an actual model prepared. It is to be noted, that the testing of ESP-32 CAM module had been carried out independently with an FTDI module, and the CAM shares no physical connection with the Arduino board. A prototype has been prepared to analyse the real time performance of the system. The front view and the internal connections of the prototype are shown in Fig.4 and Fig.5 respectively.
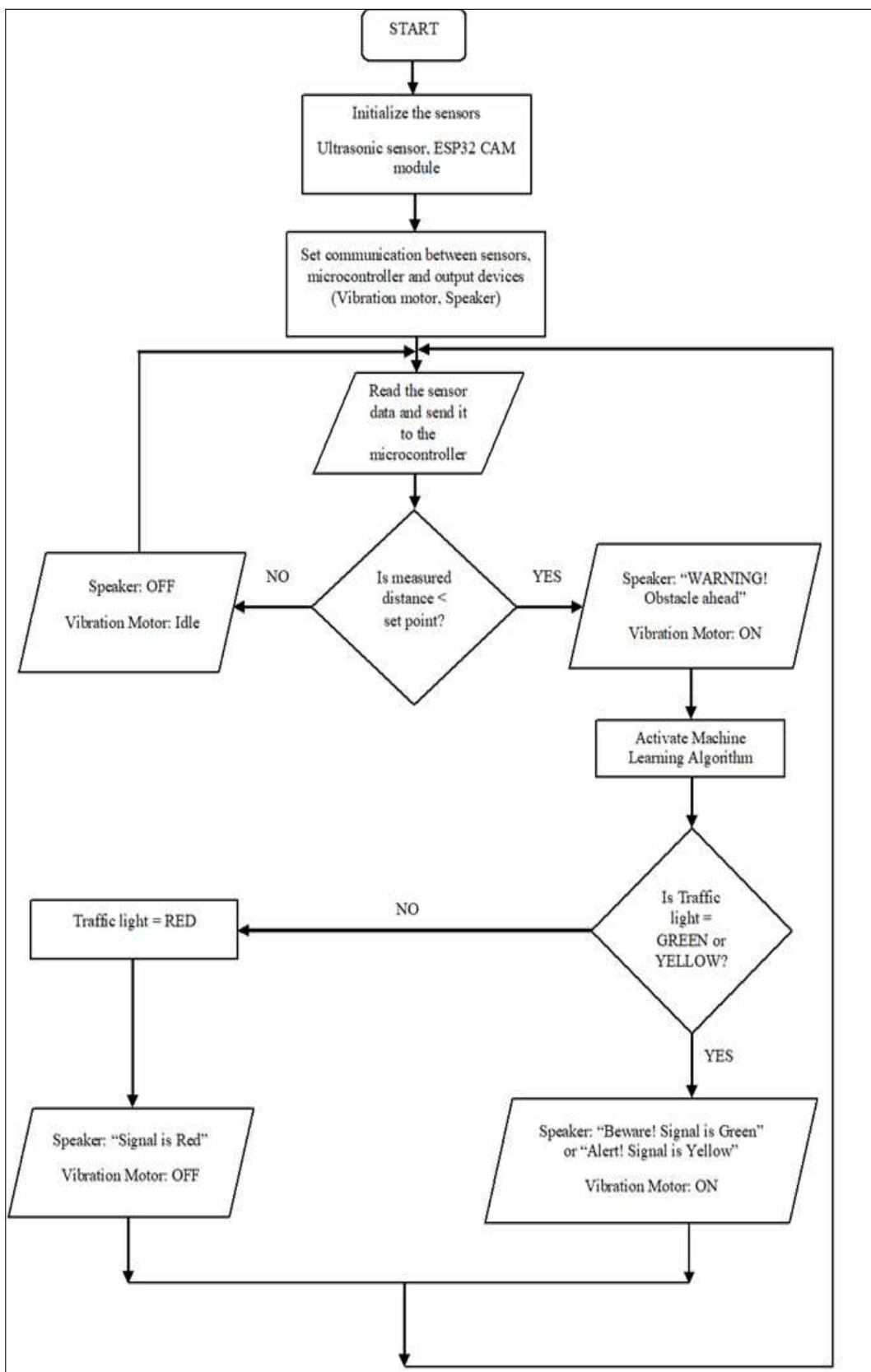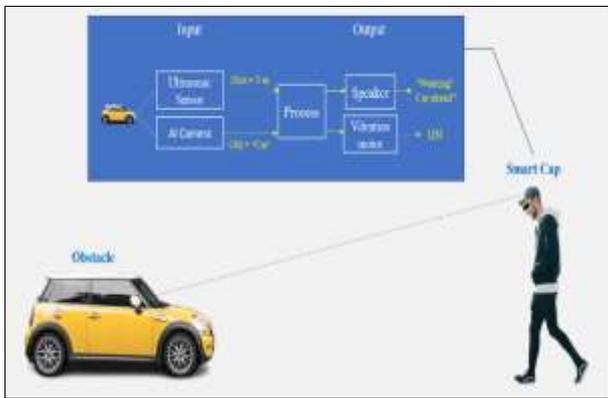
Fig.1. Smart Cap Flow Chart
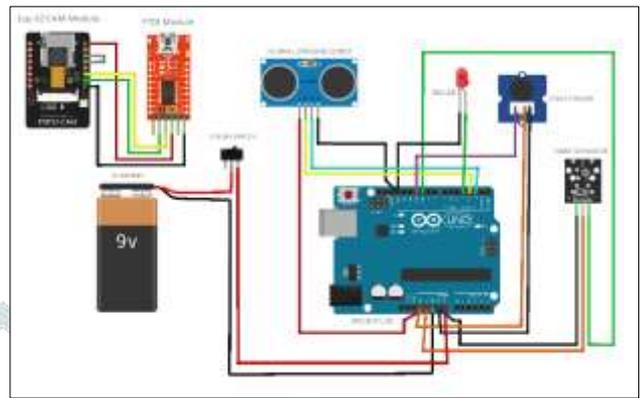
Fig.2. Workflow schematic

Fig.3. Interconnections of modules with Arduino and FTDI
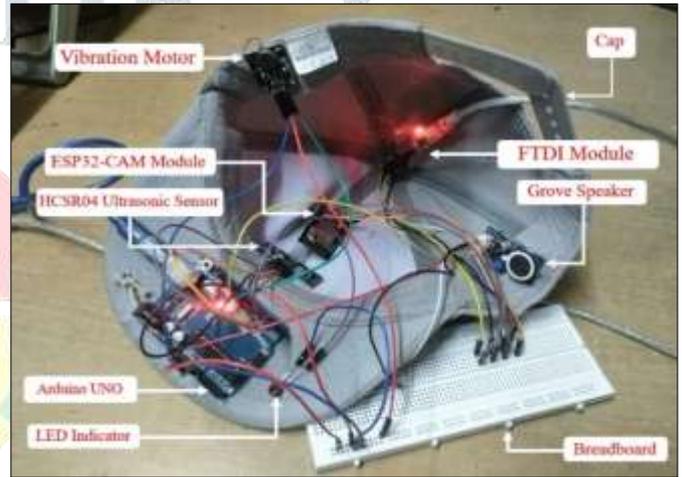
.

Fig.4. Front view of the Smart Cap prototype

Fig.5. Inside view of Smart Cap prototype

III. EXPERIMENTAL SET-UP AND RESULT ANALYSIS

We carried out numerous laboratory trials for verifying the effectiveness of our prototype. The main objective of these experiments was to validate the individual as well as combined efficiencies of ultrasonic sensor, camera module & its AI counterpart, speaker and vibration motor. The actual distance between the sensor and an object was estimated by a tape measure and the sensor output was compared with the measured. Table I summarizes the experimental results.

TABLE I DISTANCE MEASUREMENT USING ULTRASONIC SENSOR

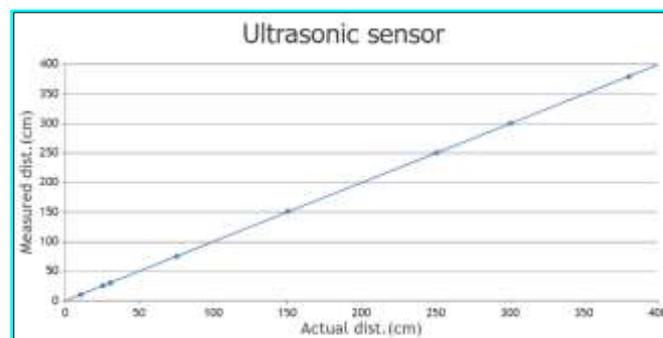| Test case	Actual distance (cm)	Sensor output (cm)
1.	10.5	10.3
2.	25.5	25.5
3.	30.5	30.0
4.	75.5	75.2
5.	150.5	151.3
6.	250.5	250.5
7.	300.5	300.0
8.	380.5	379.1

Fig.6. Characteristic curve of ultrasonic sensor

Fig.6 illustrates a graphical representation of Table I. As seen, the characteristics of the ultrasonic sensor is fairly linear. One of the prime aspects of our system is the image recognition. A real-time object recognition algorithm, YOLO v3 has been utilized for

this purpose. COCO, a large image recognition dataset trains and evaluates YOLO in object detection. It has the capability to recognize a variety of objects in still images, pre-recorded videos or a live stream.

Our experiments focused on identifying YOLO's effectiveness in recognizing day-to-day objects that a visually and audibly impaired people may encounter. Numerous images and videos have been tested for this purpose, and it is found that the device is capable of identifying multiple objects on a single frame with varying confidence level.

The algorithm proved successful at identifying common outdoor items. These include traffic lights, pedestrians, vehicles etc. with varying degrees of confidence. The algorithm should function in collaboration with the ultrasonic sensor by filtering out all distant objects/obstacles and recognize only those within a specified range.

The ESP-32 CAM module with its AI counterpart have proven to be effective in object recognition. Our experiments with plenty of other photos and videos produced results that were consistent with this pattern. The detailed observation presented in Table II shows the remarkable percentage success rate of algorithm for individual group of objects during the trials. It is to be noted, that the recorded percentage success rate of the algorithm for the obstacles like 'Pedestrians', 'Traffic light' and 'Animals' etc. was fairly precise, which is the sole requirement for a device like Smart Cap.

TABLE II YOLO v3 TEST RESULTS

Sl. No.	Objects	% Success Rate
1.	Pedestrians	98.7
2.	Vehicles (cars, motorbikes, trucks etc.)	93
3.	Traffic light	98.1
4.	Animals (dogs, cats etc.)	100
5.	Random objects (backpacks, plants etc.)	97

Overall, an astonishing 97.36% average accuracy has been recorded during these examinations.

$$\text{Average \% Success Rate} = \frac{\sum(\% \, Success \, Rate)}{5} = 97.36$$

The output modules (vibration module and Grove speaker) responded as anticipated during these tests. A prompt response with a minimum latency was recorded during the laboratory trials. Appropriate execution of codes and the on-board settings provided a timely and clear indication via audio and vibration.

By considering the individual performance of each module, and the combination of them as a system, we can conclude that the Smart Cap is an effective, advanced, yet easy to use device for its intended users. However, as indicated earlier, the study of image recognition was done independent of the Arduino due to its inability to support internet, and the integration of AI Cam with Arduino remains subject to further studies. Raspberry Pi or other cutting-edge microcontroller boards can easily take the place of Arduino-UNO to resolve this problem.

IV. CONCLUSION

The advanced functionality of Smart Cap makes it an all-in-one solution for the visually and audibly impaired individuals. The prototype's experimental analysis yielded promising results and quantified its efficiency. YOLO v3 was efficient in recognizing objects, and the recognition examination recorded an overall success rate above 97%. The Smart Cap is efficient, low-cost (25USD approx.), lightweight, easy to operate and is capable of generating sound, as well as vibration alerts, making it a good fit for both visually and audibly impaired people. Thus, our Smart Cap is ascertained to be a valuable belonging for its intended users.

The device is, however, in its probationary stage, and needs a great deal of further modifications. For the further research, it is essential to analyze the limitations of the proposed device-

1) An advanced microcontroller board such as Raspberry Pi is required to carry out seamless operations of such complex system due to the limitations of Arduino UNO in terms of processing power, memory, internet support etc.
2) A $360°$ coverage of the surrounding is an essential aspect of this type of devices. Currently, we are working on this specific problem.
3) The AI CAM requires an uninterrupted Wi-Fi connectivity to function. This renders the device useless in no/poor network coverage areas. The issue will be resolved by integrating a GSM module.
4) The Smart Cap is presently incapable of 'reading'. So, a text scanner, once successfully developed, will be incorporated into our Smart Cap.

CONFLICTs OF INTEREST

The authors declare that they have no conflicts of interest

REFERENCES

[1] "Blindness and vision impairment", who.int [online], Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 09-Oct-2022].
[2] "Vision Is Our Dominant Sense", brainline.org [online], Available: https://www.brainline.org/article/vision-our-dominant-sense. [Accessed: 22-Oct-2022].
[3] S. Kulkarni, C. Gilbert, N. Giri, P. Hankare, K. Dole, and M. Deshpande, "Visual impairment and blindness among children from schools for the blind in Maharashtra state, India: Changing trends over the last decade," *Indian Journal of Ophthalmology*, vol.70, no.2, p. 597, 2022, doi: 10.4103/ijo.ijo_1930_21.

[4] B. Li, J. B. Munoz, X. Rong, Q. Chen, J. Xiao and Y. Tian., "Vision-Based Mobile Indoor Assistive Navigation Aid for Blind People," in IEEE Transactions on Mobile Computing, vol.18, no.3, pp.702-714, 1 March 2019, doi: 10.1109/TMC.2018.2842751.

[5] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Virtual-Blind-Road Following-Based Wearable Navigation Device for Blind People", IEEE Transactions on Consumer Electronics, vol. 64, no. 1, February 2018, DOI: 10.1109/TCE.2018.2812498

[6] E. Cardillo, V. D. Mattia, G. Manfredi, P. Russo, A. D. Alfredo De Leo, A. Caddemi, and G. Cerri, "An Electromagnetic Sensor Prototype to Assist Visually Impaired and Blind People in Autonomous Walking" IEEE Sensors Journal, Vol. 18, No. 6, March 15, 2018, pp.2568-2576, doi 10.1109/JSEN.2018.2795046

[7] K. Conley, A. Foyer, P. Hara, T. Janik, J. Reichard, J. D'Souza, C. Tamma, and C. Ababei, "Vibration Alert Bracelet for Notification of the Visually and Hearing Impaired" Journal of Open Hardware, 3(1):4, pp.1–11, DOI: https://doi.org/10.5334/joh.17

[8] M. A. Khan, P. Paul, M. Rashid, M. Hossain and M. A. R. Ahad, "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind," in IEEE Transactions on Human-Machine Systems, vol. 50, no. 6, pp.507-517, Dec. 2020, doi: 10.1109/THMS.2020.3027534.

[9] S. A. Satpute, J. R. Canady, R. L. Klatzky and G. D. Stetten, "FingerSight: A Vibrotactile Wearable Ring for Assistance With Locating and Reaching Objects in Peripersonal Space," in IEEE Transactions on Haptics, vol.13, no.2, pp. 325-333, 1 April-June 2020, doi: 10.1109/TOH.2019.2945561.

[10] S. Qiu, J. Hu, T. Han, H. Osawa, and M. Rauterberg, "An Evaluation of a Wearable Assistive Device for Augmenting Social Interactions", IEEE Access, Vol. 8, 2020, pp.164661-164677, DOI: 10.1109/ACCESS.2020.3022425.

[11] A. S. Mahdi, A. H. Muhsin, A. I. Al-Mosawi, "Using ultrasonic sensor for blind and deaf persons combines voice alert and vibration properties. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.13633415.v1

[12] S. Nazim, S. Firdous, S. R. Pillai and V. K. Shukla, "Smart Glasses: A Visual Assistant for the Blind," 2022 International Mobile and Embedded Technology Conference (MECON), 2022, pp. 621-626, doi: 10.1109/MECON53876.2022.9751975.

[13] K. Xia, X. Li, H. Liu, M. Zhou and K. Zhu, "IBGS: A Wearable Smart System to Assist Visually Challenged," in IEEE Access, vol. 10, pp. 77810-77825, 2022, doi: 10.1109/ACCESS.2022.3193097.

[14] X. Hu, A. Song, Z. Wei and H. Zeng, "Stereo Pilot: A Wearable Target Location System for Blind and Visually Impaired Using Spatial Audio Rendering," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1621-1630, 2022, doi: 10.1109/TNSRE.2022.3182661.