

# Indoor Navigation Aid System for the Visually Impaired

Tanmay Laud

Bachelor of Technology in Electronics and  
Telecommunications  
Veer mata Jijabai Technological Institute  
Mumbai, India

Shubham Jain

Bachelor of Technology in Electronics and  
Telecommunications  
Veer mata Jijabai Technological Institute  
Mumbai, India

**Abstract**—The visually impaired, in a country like India, are always dependent on others for navigation. Different factors come into play when the user navigates indoors and outdoors. In case of indoors, the environment may be known, however, several short-distance obstacles may exist (doors, staircase, table, chair, etc.), while in case of outdoors, the terrain is not known hence an external navigation aid (like Google Maps) is required. The indoor navigation device consists of a wearable headset with an ultrasonic emitter and sensors with affixed artificial pinnae. In this model, an upswept frequency-modulated (FM) ultrasound signal is emitted from a transmitter with broad directional characteristics in order to detect obstacles. The ultrasound reflections from the obstacles are picked up by a two-channel receiver. The frequency of the emitted ultrasound is swept from 35 to 50 kHz within 3 ms, so it has almost the same characteristics as the ultrasound a bat produces for echolocation. Based on the time characteristics of the reflected ultrasound wave, a pre-processed 25:1 down-converted audio sample is selected by using a microcomputer. These audible waves are then presented binaurally through earphones. In this method obstacles may be perceived as localized sound images corresponding to the direction and the size of the obstacles. This mobility was evaluated based on psychophysical experiments. Further, we employ Computer Vision and Machine Learning to detect objects like doors, stairs and windows aid via speech assistance

**Keywords**—Ultrasonic; navigation; healthcare; biomedical; signal processing; echolocation; Raspberry Pi; blind mobility aid; Tensorflow; MobileNet;

## I. THE PROPOSED DEVICE

We present a device that combines principles of ultrasonic echolocation and spatial hearing to provide human users with environmental cues that are 1) not otherwise available to the human auditory system and 2) richer in object, and spatial information than the more heavily processed sonar cues of other assistive devices. A device that uses a forehead-mounted speaker to emit ultrasonic “chirps” (FM sweeps) modeled after bat echolocation calls. The echoes are recorded by bilaterally mounted ultrasonic microphones, each mounted inside an artificial pinna, also modeled after bat pinnae to produce direction-dependent spectral cues. After each chirp, the recorded chirp and simulated reflection are played back to the user. This magnifies all temporally based cues linearly by a factor of  $m$  and lowers frequencies into the human audible range. For empirical results reported here,  $m$  is 20 or 25 as indicated. That is, cues that are normally too high or too fast for the listener to use are brought into the usable range simply by replaying them more slowly.

## II. PROBLEM STATEMENT

To provide a composite reliable solution for navigation in indoor and outdoor circumstances for visually challenged people.

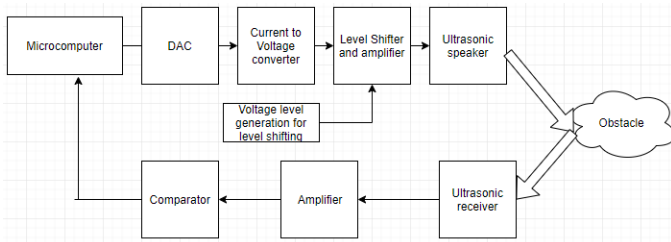
*Elaboration:*

In the pre-internet days, there was no way for anyone to travel from one place to another without asking for directions from some other person who was well-versed with the directions. However, the times have changed as nowadays, it is even possible for a child to walk from school to home just by using a smart phone. Google has changed the way people used to think. With the advent of Google Maps service, it is possible for the development of many other services that are related to maps. For example, Ola, Uber, Zomato are the heavily used apps by this generation and they make use of the services provided by Google Maps for location detection and navigation. However, with these developments taking place, the people who are visually challenged are being left out. The focus needs to be on how we can use these services so that they can be beneficial to the people who are challenged in one or the other way. These services should provide great user experience and make the life of the visually challenged people simpler. The GPS based navigation services, however, face a drawback. They are not accurate enough to be used indoors, that is, inside buildings. Google Maps cannot be used to go from one door to another. Hence, a comprehensive indoor navigation solution is required for the visually impaired, that runs in co-operation with the outdoor navigation service. Although, a number of electronic travel aids that utilize sonar have been developed, none appear to be in common use, and very few provide information other than range finding or a processed localization cue. For example, distance to a single object is calculated and, then, mapped to a sound frequency, providing only extremely limited information about the world. Cost of the device is another major issue that is often neglected. Therefore, a cost-effective composite solution for indoor as well as outdoor navigation is the need of the hour.

## III. SYSTEM DESIGN

To use echolocation artificially for navigation and obstacle detection, a device needs to be designed which transmits a signal and makes deductions about environment characteristics from the reflected signals.

The DC power supply used in the circuit is +12 V and -12 V, which is very standard, and can be taken from lead acid batteries or adapters.

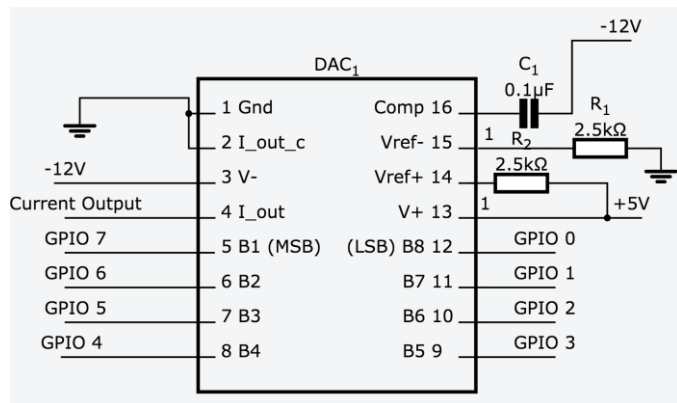


**A. Selection of Microcomputer**

The microcomputer used is a Raspberry Pi (RPI) which contains Broadcom BCM2387 1.2 GHz Quad-Core Cortex processor, whose GPIO pins are used to transmit the real time chirp sample values parallelly. The chirp consists of a logarithmic upward sweep from 35 to 50 kHz. The chirp samples are fed real time to GPIO pins rather than transferring pre-calculated values, to account for distortion which may be caused due to internal processes occurring in Raspberry Pi. The library which could support high speed data transfer is needed in Raspberry Pi. Wiring-Pi is the library which supports this high speed GPIO transfer in C programming. So, this library is used. The chirp is created for a period of 3 ms.

**B. Digital to Analog Conversion**

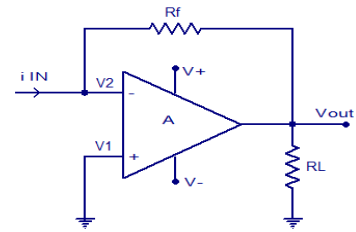
The DAC which is to be used to convert this parallel data into serial data needs to operate at a high sampling rate of at least 400 ksp/s, which is GPIO transfer speed. DAC0808 satisfies these requirements. The connections for DAC0808 are as follows.



The high output on GPIO pins gives 3.3 V from Raspberry Pi, which is high logic for DAC0808, so no voltage conversion is required. The reference voltage used here is 5 V. The DAC gives current output which is 2 mA for digital input of 255(maximum input). So, the resistance values used are 2.5 K since  $(2.5 K) * (2 mA) = 5V$ . The DAC standard supplies are +5V and -15 V.

**C. Current to Voltage Conversion**

The current output needs to be converted into voltage output. This is done with the help of a current to voltage converter. This is built with the help of an op amp. The op-amp too needs to operate at very high speed. LF356 is an op amp with a very high slew rate of 16 V/us, so it can be used as a current to voltage converter, as it can vary its output at a very high speed.

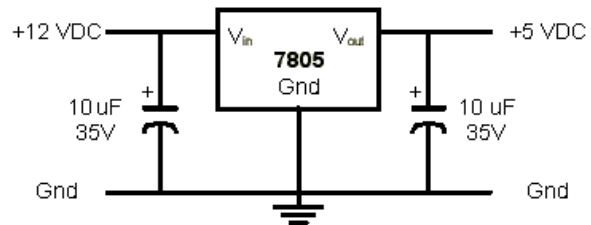


Active Transimpedance amplifier / current to voltage converter

The  $R_f$  value to be used for this circuit diagram is 2.5K, as explained above. The pin diagram of LF356 is same as general purpose conventional op amp LM741.

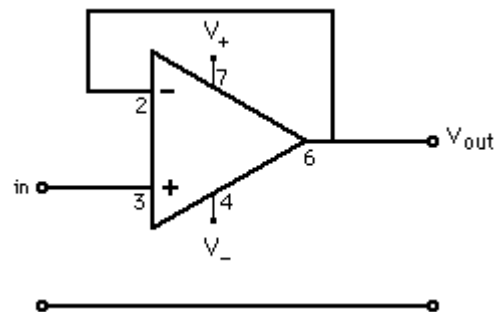
**D. Voltage Level Shift**

The output of the current to voltage converter lies between 0 to 5 V. The speaker MA40SRS which we are using can operate from -20 V to 20 V peak to peak, however due to power supply constraints, we restrict ourselves to -10 V to 10 V. For this, first the DC level has to be brought to zero and then amplification has to be done. The DC level at the output is 2.5 V. To create this reference, LM7805 is used to generate 5V from 12V and then, output is fed to a voltage divider, with equal resistors to create 2.5 V reference. The connections for 7805 are as follows.



The output of this 7805 is connected to a voltage divider to create 2.5 V reference.

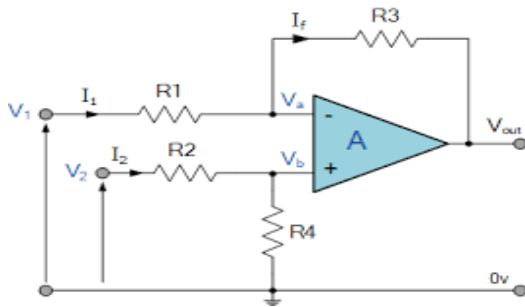
We need 2.5 V reference for subtractor amplifier, which might sink current reducing the threshold voltage of 2.5 V, so a buffer is needed which will ensure no current is sinking. The buffer is made up of an op amp which has very high input impedance, and very low output impedance which is as required. The buffer design is as follows



The Operation amplifier here can be either LF356 or LM741.

**E. Difference Amplification**

The Subtractor amplifier is used to bring the DC level to 0 and then make the output to -10 to 10 V peak to peak, with a gain of 4(as the input after subtraction will be -2.5 to 2.5 V peak to peak).



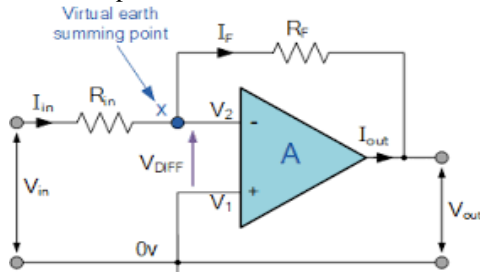
At inverting terminal, the input is the threshold 2.5 V and non-inverting input is the output of current to voltage converter. The equation governing output of subtraction amplifier is

$$V_{out} = \frac{R_3}{R_1} (V_2 - V_1) \text{ given } R_3=R_4 \text{ and } R_1=R_2$$

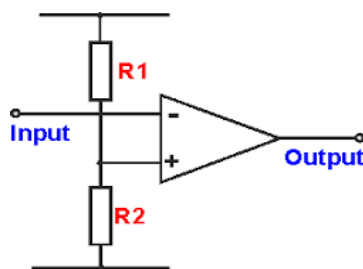
So,  $R_3=20K$  and  $R_1 = 5K$  will satisfy the condition of gain 4 and subtraction. The output is the fed to the transmitter MA40S4S.

**F. Receiver Design**

The microcomputer is programmed to transmit the chirp for a period of 3 ms and receive the reflected signal for a period of another 27 ms. The receiver MA40S4R output is fed to an inverting amplifier as the received signal is in millivolts. The gain of the amplifier is 100.



An  $R_f$  value of 100k and  $R_{in}$  value of 1 K will suffice our requirements. This output is then fed to comparator which can be used to provide a trigger to the raspberry pi if the input exceeds a threshold, indication a signal has been received. The comparator threshold is set using trial and error and the voltage level is created with the help of a voltage divider with one resistor fixed and other resistor in form of 10K pot which can be varied to vary the threshold. The comparator can be either LF356 or LM741.



$R_1$  is fixed at 1K and  $R_2$  is 10K variable pot. There are two receivers, so two trigger points for left and right detection are connected to 2 RPi GPIO pins.

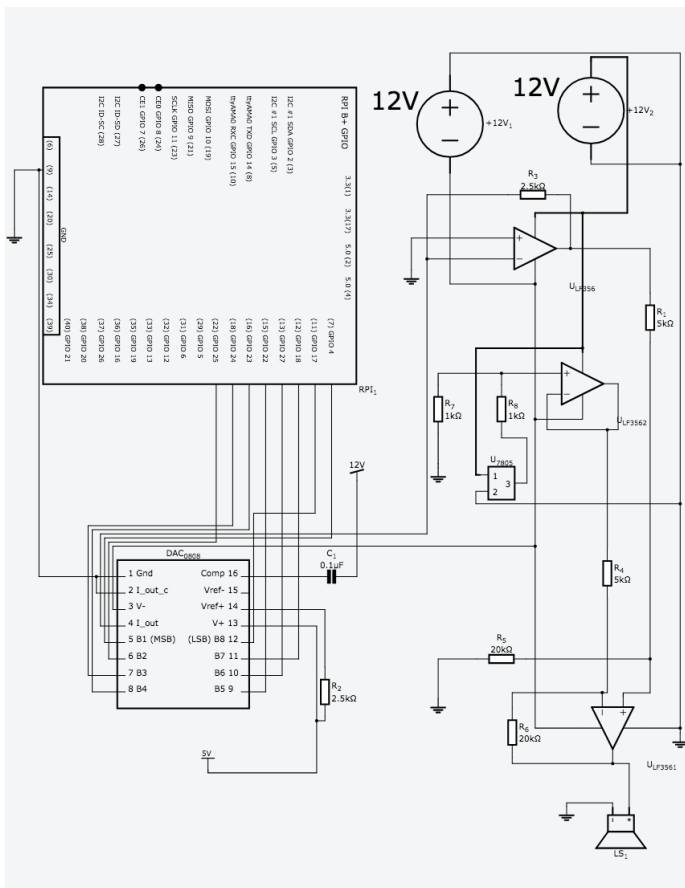
**G. Audio Playback Generation**

The RPi is programmed to store values for triggers from 8 obstacles on both left and right channels. The time values for a particular obstacle are then used to decide which wav file to be played to create the impression about location of that obstacle for the subject. First the absolute value of time gives information about the distance of the obstacle from the subject, and the time at which the echo signal is to be played. For example, if the trigger time is in upper range of 27 ms, the amplitude of the echo signal is kept low. Further the difference between the left and right trigger times for an obstacle, give indication about the left-right localization of the obstacle. For example, if the obstacle is very close to the left receiver, the left trigger time will be sufficiently less than right trigger time. The decision for which wav files to be played for which obstacle and at what times is made as explained above. Then, the original sweep signal is down-converted in range of 1400-2000 Hz and played first, the time period of this wav file is 75 ms=25\*3ms. Then the wav files for each of the obstacles is played over period of 750 ms. Then, a break of 1250 ms is taken for all the residual reflections from previous chirp to vanish, and then the sweep is transmitted once again.

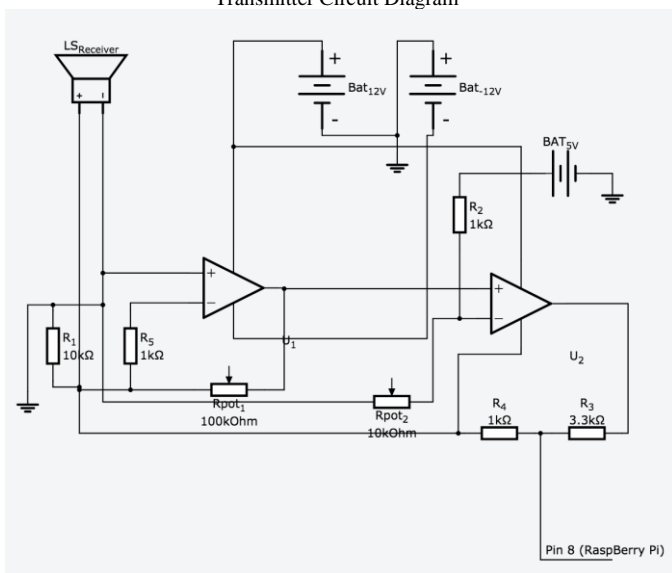
**III. SYSTEM IMPLEMENTATION**

**A) Echolocation Device**

The basic idea was to create a device that uses a forehead-mounted speaker to emit ultrasonic “chirps” (FM sweeps) modelled after bat echolocation calls. The echoes are recorded by bilaterally mounted ultrasonic microphones, each mounted inside an artificial pinna, also modelled after bat pinnae to produce direction-dependent spectral cues. After each chirp, the recorded chirp and reflections are played back to the user at 1/m of normal speed, where m is an adjustable magnification factor. This magnifies all temporally based cues linearly by a factor of m and lowers frequencies into the human audible range. There are certainly some devices developed using this approach, but none of them is cost effective as most of them use laptops and similar devices for software functionality, expensive sound cards for digital to analog conversion and highly expensive ultrasonic speaker and microphones. We have tried to improve cost effectiveness by using comparatively cheaper sets of hardware components. As a microcomputer, we used a Raspberry Pi 3 model B which could perform the required software functionalities. We used DAC0808 for digital to analog conversion and general-purpose op-amps for amplification and MA40S4S and MA40S4R for transmission and reception respectively. The current technique uses transmitter and receiver circuits as shown in figures below.



Transmitter Circuit Diagram



Receiver Circuit Diagram

- 1) The transmitter uses 8 pins on Raspberry Pi as GPIO to transmit parallel real-time sample values for ultrasonic chirp in Range of 35 to 50 KHz. The range is decreased considering the stringent frequency characteristics of transmitter and receiver.
- 2) The DAC0808, takes these values and converts them into analog output, which is in range of 0-5 V since the  $V_{ref}$  given to DAC0808 is 5 V.
- 3) The output capacity for speaker is -20 to 20V peak to peak, however due to commercial power supply of

12V available, we use -10 to 10 V peak to peak for transmission. The output of DAC is first converted to voltage output, as DAC0808 is current output with the help of LF356 op amp, which has a very high slew rate to operate at high frequencies.

- 4) The output is the subtracted from 2.5 V to shift the dc level to 0, and then this difference is amplified with gain 4 to give 20 V peak to peak output.
- 5) The receiver circuit first amplifies the incoming signal which is in millivolts with a gain of 100 using op amp LF356.
- 6) Then this output is fed to a comparator comparing the input with a standard threshold to detect if a signal is received at a particular time, if the input exceed a threshold, which is set by trial and testing, then received signal is detected.
- 7) With the help of software in RPi, 10 such obstacle time periods are detected for both left and right microphones.
- 8) Then, depending upon the absolute times and relative times between left and right interrupt time values, the wav files to be played are selected from the list of wav files available.
- 9) The list is formed through trial and error, for e.g. by testing the conditions for an obstacle very close to left ear, a wav file is formed which has a very high amplitude on left ear and so on.
- 10) The original transmitted sweep is down-converted and played, and the wav files selected with the above process are played as echoes, to give a sense that the actual echo signal is played, which was the actual aim as mentioned above.

As explained earlier, based on the time characteristics of the reflected impulse, we determine the approximate distance of the object. Then, the corresponding audio echo sample is played binaurally via earphones. For example, if the obstacle is approximately within one meter range and oriented towards the left of the user plane, then the stereo audio sample that is left panned is played after the original audio pulse. Thus, a left inclined echo effect is heard by the user.

### B) Object Detection Using Mobile Camera

#### 1) Training Set Acquisition

Before you start any training, you'll need a set of images to teach the model about the new classes you want to recognize. We created a dataset of photos of **doors, windows and staircase**, to use initially, 500 images of each object.

#### 2) Selection of Neural Network

MobileNet(MN) is a small efficient convolutional neural network. "Convolutional" just means that the same calculations are performed at each location in the image. They are coined "mobile-first" in that they're architected from the ground up to be resource-friendly and run quickly, right on your phone.

The main difference between the MN architecture and a "traditional" Convolutional Neural Networks is that, instead of a single 3x3 convolution layer followed by Batch normalization and ReLU, MN split the convolution into a 3x3 depth-wise convolution and a 1x1 pointwise convolution. The details of why this is so significant can be found in the *MobileNet* paper.

#### 3) Accuracy Factor:

MobileNets are not usually as accurate as the bigger, more resource-intensive networks like ImageNet. But finding that resource/accuracy trade-off is where MobileNets really shine. MobileNets surface two parameters that we can tune to fit the resource/accuracy trade-off of our exact problem: width multiplier and resolution multiplier. The width multiplier allows us to thin the network, while the resolution multiplier changes the input dimensions of the image, reducing the internal representation at every layer. Google open-sourced the MobileNet architecture and released 16 ImageNet checkpoints, each corresponding to a different parameter configuration.

The MobileNet is configurable in two ways:

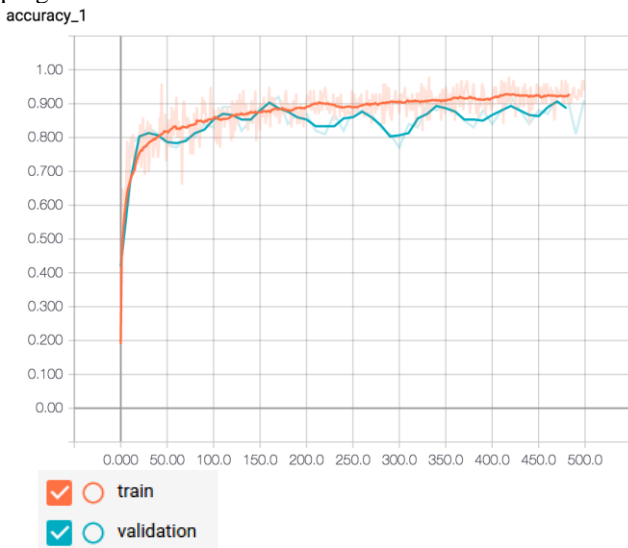
- Input image resolution: 128,160,192, or 224px. Unsurprisingly, feeding in a higher resolution image takes more processing time, but results in better classification accuracy.
- The relative size of the model as a fraction of the largest MobileNet: 1.0, 0.75, 0.50, or 0.25.

We used (224 ,0.5) for our application.

#### 4) Training Process

MobileNet does not include any of the obstacles we are training on here. However, the kinds of information that make it possible for MobileNet to differentiate among 1,000 classes are also useful for distinguishing other objects. By using this pre-trained network, we are using that information as input to the final classification layer that distinguishes our 'door,' 'window' and 'staircase' classes.

The first figure shows accuracy as a function of training progress



Two lines are shown. The orange line shows the accuracy of the model on the training data. While the blue line shows the accuracy on the test set (which was not used for training).

### IV. RESULTS

#### 1) Ultrasonic Echolocation device results:

The device was tested with 5 people blindfolded in the lab with obstacles placed at varied distances. It was observed that the users initially had a problem understanding the audio cues based on obstacles. They were able to interpret the echoes and the amplitudes differences in the left and right ears based on

the position of the obstacles placed in front of them but the distance of the obstacle had to be less than 1 meter. However, on repeated use of the device, the users were able to avoid certain large objects placed in front of them.

#### 2) Computer Vision based Obstacle Detection App results



The 3 images above are snapshots of the Classifier in live action. It takes real-time camera input from the phone's camera. The classifier then outputs the probability of the object detected. As mentioned earlier, only 3 objects have been considered, that is, doors, windows and staircase (images above show 3 samples from the trained model running on Xiaomi Redmi Note 3 mobile device, compiled on Android Studio with TensorFlow Lite API). The above accuracy measures are only valid for certain dry runs. The mean accuracy will differ depending on the dataset, the model architecture and the choice of hyper-parameters.

### V. CONCLUSION AND FUTURE SCOPE

The prototype we aimed to develop was based on the fact that the raw echoes coming from the transmission of an ultrasonic frequency chirp for a short period of time will not only contain information about the distance of the obstacle but also horizontal location information and vertical location information, as well as texture, geometrical and material cues. So, the slowed down received signal would serve as the perfect input for creating an image of the surroundings for the subject. However, as mentioned earlier, the design of the model had to be changed due to some constraints.

The new prototype was not based upon the principle of playing the raw echoes directly, but extracting features from the environment, and using those features for playing artificially made echoes for the subject. Though, after testing conditions for all types of obstacles, and including artificial echoes for all of them, this device could never be as capable as the one with raw echoes. The constraints were based on usage of high speed and low-cost ADCs which could convert the raw data coming from microphone to digital data, and feed it to the RPi and it proved to be difficult to find such ADCs. With availability of such low cost and high-speed ADCs, the current model could be extended to fulfill our dream of developing such a device. Further, there were constraints related to ultrasonic speaker and microphone. The speaker- microphone pair we used was

MA40S4S-MA40S4R, which are low cost components. However, the bandwidth and range of this pair is very restricted. The directivity in other directions is very low, as its transmission characteristics only favor the axial direction. Because of the frequency characteristics certain frequency components in the chirp to be highly diminished, and furthermore, the range is very compromised even if its maximum gain frequency is transmitted. Low directivity in other directions only enable the obstacles more or less in the axial direction to be detected. With availability of low-cost microphones and speakers with suitable characteristics as mentioned above, a better prototype for indoor navigation could be developed.

Further, the neural network based obstacle detection system can be further tested in various indoor environments along with regional speech support. The three solutions must be incorporated into one All-in-One system that doesn't require too many devices. Portability of the entire system is another area that needs to be explored in depth. The current system utilizes two 12V power supplies and providing such a supply on the go is an issue. Hence, reducing the power requirement of the system and trying to improve the range performance of the system is a challenge and builds scope for future research.

#### ACKNOWLEDGMENT

We would like to thank our Guide Dr. Faruk Kazi for guiding us throughout this project and giving it the present shape. It would have been impossible to begin the project without their support, valuable suggestions, constructive criticism, encouragement and guidance. We would also like to thank CoE-CNDS (Centre of Excellence Complex and Non-Linear Dynamic System) lab for providing the access to experimental setup developed by L&T Electrical and Automation. We also acknowledge the support of 1Step CSR initiative of Larsen and Toubro Infotech (LTI), as without them, it would not be possible to conduct the research.

#### REFERENCES

- [1] Viswanathan, K. and Sengupta, S., 2015, September. Blind navigation proposal using SONAR. In *Bombay Section Symposium (IBSS), 2015 IEEE* (pp. 1-6). IEEE.
- [2] Gonnot, T. and Saniie, J., 2016, May. Integrated machine vision and communication system for blind navigation and guidance. In *Electro Information Technology (EIT), 2016 IEEE International Conference on* (pp. 0187-0191). IEEE.
- [3] Ma, J. and Zheng, J., 2017, June. High precision blind navigation system based on haptic and spatial cognition. In *Image, Vision and Computing (ICIVC), 2017 2nd International Conference on* (pp. 956-959). IEEE.
- [4] Ali, A. and Ali, M.A., 2017, October. Blind navigation system for visually impaired using windowing-based mean on Microsoft Kinect camera. In *Advances in Biomedical Engineering (ICABME), 2017 Fourth International Conference on* (pp. 1-4). IEEE.
- [5] Patel, K.K. and Vij, S.K., 2008, April. Externalizing virtually perceived spatial cognitive maps. In *Systems Conference, 2008 2nd Annual IEEE* (pp. 1-7). IEEE.
- [6] Ifukube, T., Sasaki, T. and Peng, C., 1991. A blind mobility aid modeled after echolocation of bats. *IEEE Transactions on biomedical engineering*, 38(5), pp.461-465
- [7] Riehle, T.H., Lichter, P. and Giudice, N.A., 2008, August. An indoor navigation system to support the visually impaired. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* (pp. 4435-4438). IEEE.
- [8] Strumillo, P., 2010, May. Electronic interfaces aiding the visually impaired in environmental access, mobility and navigation. In *Human system interactions (HSI), 2010 3rd conference on* (pp. 17-24). IEEE..
- [9] Kasthuri, R., Nivetha, B., Shabana, S., Veluchamy, M. and Sivakumar, S., 2017, March. Smart device for visually impaired people. In *Science Technology Engineering & Management (ICONSTEM), 2017 Third International Conference on* (pp. 54-59). IEEE.
- [10] Chern, A., Lai, Y.H., Chang, Y.P., Tsao, Y., Chang, R.Y. and Chang, H.W., 2017. A smartphone-based multi-functional hearing assistive system to facilitate speech recognition in the classroom. *IEEE Access*, 5, pp.10339-10351.
- [11] Endo, Y., Sato, K., Yamashita, A. and Matsubayashi, K., 2017, September. Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM. In *Biometrics and Kansei Engineering (ICBAKE), 2017 International Conference on* (pp. 158-162). IEEE.
- [12] Jiang, H., Gonnot, T., Yi, W.J. and Saniie, J., Computer Vision and Text Recognition for Assisting Visually Impaired People using Android Smartphone.
- [13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017