Scientific
Research
Publishing

Enhancing the Efficiency of Voice Controlled Wheelchairs Using NAM for Recognizing Partial Speech in Tamil

Angappan Kumaresan[1,2]*, Nagarajan Mohankumar[3], Mathavan Sureshanand[4], Jothi Suganya[5]

[1]Anna University, Chennai, India
[2]Department of Computer Science and Engineering, SKP Engineering College, Tiruvannamalai, India
[3]Department of Electronics and Communication Engineering, SKP Engineering College, Tiruvannamalai, India
[4]Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India
[5]Department of Electronics and Communication Engineering, Mailam Engineering College, Villupuram, India
 Email: *kummaresan@gmail.com

Abstract

In this paper, we have presented an effective method for recognizing partial speech with the help of Non Audible Murmur (NAM) microphone which is robust against noise. NAM is a kind of soft murmur that is so weak that even people nearby the speaker cannot hear it. We can recognize this NAM from the mastoid of humans. It can be detected only with the help of a special type of microphone termed as NAM microphone. We can use this approach for impaired people who can hear sound but can speak only partial words (semi-mute) or incomplete words. We can record and recognize partial speech using NAM microphone. This approach can be used to solve problems for paralysed people who use voice controlled wheelchair which helps them to move around without the help of others. The present voice controlled wheelchair systems can recognize only fully spoken words and can't recognise words spoken by semi-mute or partially speech impaired people. Further it uses normal microphone which has severe degradation and external noise influence when used for recognizing partial speech inputs from impaired people. To overcome this problem, we can use NAM microphone along with Tamil Speech Recognition Engine (TSRE) to improve the accuracy of the results. The proposed method was designed and implemented in a wheelchair like model using Arduino microcontroller kit. Experimental results have shown that 80% accuracy can be obtained in this method and also proved that recognizing partially spoken words using NAM microphone was much efficient compared to the normal microphone.

---

*Corresponding author.

Keywords

NAM, Speech Recognition, TSRE, Wheelchair Guidance, HCI

1. Introduction

The ordinary human speech, which is conducted through air, is produced by the vocal cord vibration due to the air stream from lungs. Talking, singing, whispering, crying, laughing are other different types of voices apart from ordinary speech. Voice conduction through bone is the reason that enables us to hear our own voice whereas others' voice can be heard after it is conducted through air. This is the major reason that people hear their own voice sound different when heard from a recording. Voice can be classified into two types. One of the types is produced as a result of air stream passing through the glottis and is termed as phonation sound. Another type is termed as conduction sound which is produced as a result of the vibration of vocal tract wall due to the rapid vibration of air. The above specified sound passes through the tissues and bones which can be sensed and recognized using body conduction microphone. This is a soft murmur which cannot be heard by people around [1]. Nakajima *et al.* introduced the microphone based on the concept of stethoscope. A stethoscopic microphone is used to recognize the soft speech that is conducted through the bone and muscles. Thus the soft murmur passing through the bone and muscle is called Non Audible Murmur (NAM). Using NAM microphone we can hear very quietly uttered speech without noise by making use of the soft silicon layer attached to the human mastoid part which is applicable for personal conversation in public places [2]. NAM is also applicable for impaired people who have problem with vocal cords. Human Computer Interaction is gaining importance these days and speech is an appropriate choice for bridging the gap between humans and machines. Though wheelchairs are available which can be controlled with the help of speech commands [3], it will not be effective in recognizing partial speech. So we are proposing to use NAM speech for guiding the wheelchair. The process is involved in controlling the wheelchair with the help of speech inputs. The process involves directing the voice of the speaker into the voice controlled wheelchair system which is in turn fed into the structure containing the combination of power supply with micro controlled circuit. The output signal from this structure is fed into the DC motor circuit for controlling the motion of the motors attached to the wheelchair for controlling the different motion of the wheelchair. So we are proposing to use NAM speech for guiding the wheelchair. The optimal position for attaching the NAM microphone is shown in **Figure 1**.

Speech recognition is an important concept that would help most of impaired people. Hui Ye *et al.* have worked on voice morphing technique which can modify the speech of a source speaker to that of the destination speaker. The main process in voice morphing technique is to match the spectral envelope of the source and target speakers along with linear transformations derived from parallel training data aligned with respect to time. The authors have studied about the linear transformation approaches and have dealt with two main issues. One of the issues was the need of parallel training data for deriving transform estimation and so a general maximum

Figure 1. NAM microphone attached to the user.

likelihood framework was designed to avoid the need of parallel training data. Secondly, several compensation techniques were developed to mitigate glottal coupling, spectral variance of unvoiced sounds which were found to be the major reasons for the artifacts [4].

Chaficmokbel introduced the clustered framework for online adaptation of Hidden Markov Model (HMM) parameters for real life scenarios and thus increased the efficiency of speech recognition systems. Some techniques have been additionally developed to control the adaptation convergence in unsupervised modes. Bayesian adaptation and spectral transformation are the two approaches that have been used for adapting HMM parameters to new conditions. The author has developed a unified framework where both of the above specified methods are considered. For each Gaussian distribution, the framework attributes one transformation and the latter is partitioned considering the adaptation data. The same parameter vector is shared among the transformation of each class. Maximum *a posteriori* (MAP) is used for determining the parameters of global transformation based on original HMM *a priori* distributions. The author has implemented the general adaptation algorithm within the CNET speech recognition system that was evaluated on several telephone databases and using these methodologies, a systematic convergence was achieved in an online unsupervised mode. In this unified approach, considering $N_1(., \mu_1. \Gamma_1)$ and $N_2(., \mu_2. \Gamma_2)$ to be two Gaussian distributions which are associated with acoustic frames such as n_1 and n_2 during training on a large database, the N_1 and N_2 can be replaced with a single Gaussian function N_3 which models the whole frames such as $n_3 = n_1 + n_2$. A loss of precision occurs as a result of the modelling process. The loss of precision might be the log likelihood ratio between the new distribution N_3 and the old distributions (N_1, N_2) on the n_3 frames which is represented in terms of the distance between the two distributions [5].

$$d\left(N_1, N_2\right) = \log \frac{\left\|\Gamma_3\right\|^{(n_1+n_2)/2}}{\left\|\Gamma_1\right\|^{n_1/2}\left\|\Gamma_2\right\|^{n_2/2}} \tag{1}$$

A robust speech recognition method based on stochastic mapping technique was proposed by Mohammed Afify *et al.*, in which a Gaussian mixture model was constructed for joint distribution of noisy and clean features. During testing, this distribution was used to predict clean speech. Two estimators such as iterative based on maximum *a posteriori* (MAP) and Minimum Mean Square Error (MMSE) were considered. The resulting estimators were a mixture of linear transforms derived from joint distribution. The combination of the proposed system with the CMLLR was found to be more efficient compared to the use of CMLLR alone in real field data [6]. Rory A. Cooper *et al.* have worked on Electrically Powered Wheelchair (EPW) guidance using isometric joysticks. The authors tried to maximize user requirements and minimize their limitations. So they have extended Fitts' law to support target acquisition even in the case of a continuously updated target. EPW driving was examined with the help of this extended law with the help of a standard position sensing joystick along with a prototype Isometric Joystick (IJ). Significant differences were observed while using these two types of joysticks in terms of movement time, root mean square error, movement time and other factors during turning movements. The authors have verified that IJ provides better turning performance comparable with others [7]. Apart from this, Qiang Zeng *et al.* have worked on Collaborative Wheelchair Assistant (CWA) which is entirely dependent upon the motion planning skills of the user for manoeuvrings the wheelchair. The user could decode the path and control the speed whereas the software helps the wheelchair to move in a defined guide path. A path editor was available which helped the user to modify the path of the wheelchair in case of any obstacles. This system is found to be more effective in terms of performance and cost without any need of complex sensors or artificial intelligence decisions [8].

2. Speech Recognition in Wheelchairs

In existing methodologies, locomotion is not fit for impaired people where wheelchairs play a major role in transporting them from one place to another. Since speech plays a major role in controlling the movement of wheelchairs, we should understand the nature and the waveform of normal speech in order to understand the detailed structure of normal speech which is shown in **Figure 2**.

Mohammed Fezeri explained about voice commands for controlling the wheelchair and sensors to detect and avoid obstacles during motion. The author have used hybrid method, in which four ultrasonic sensor modules are used to find the obstacles and avoid them along with a set of classical voice recognition commands facilitating the wheelchair movement. For real time experiment, the microcontroller gives the environment information
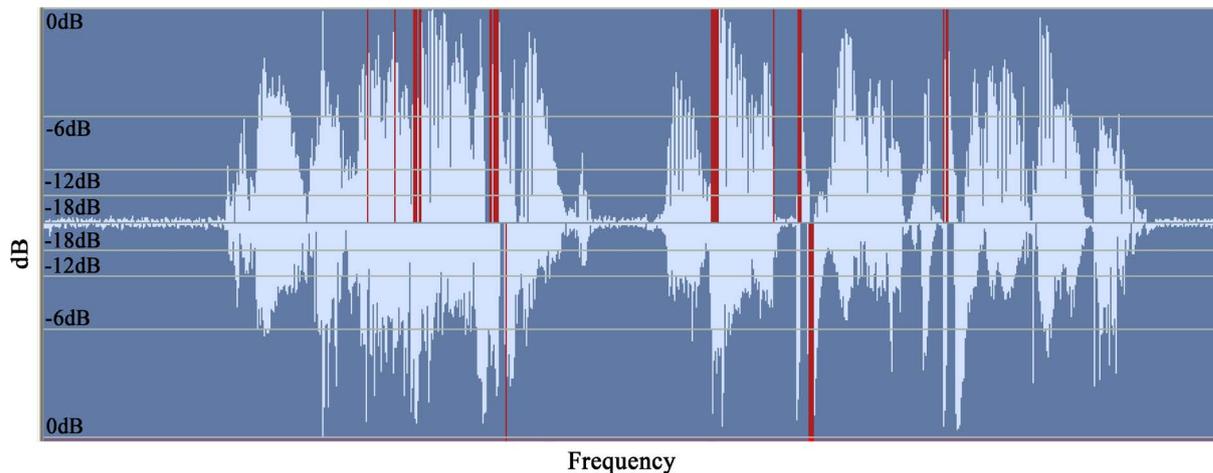
Figure 2. Waveform of normal speech.

to the PC which recognize the voice commands and makes decision whether the command is to be executed or not. The main purpose of sensors is to enable and disable the wheelchair when the obstacles are observed. Artificial intelligence concepts are developing in a rapid rate and are gaining importance in our day to day life which leads to AGV (Automatic Guided Vehicle) navigation for movement of wheelchair guidance with speech recognition. The hybrid method is simple and effective to be used for impaired people. For wheelchair movements, ten Arabic words were used in AVG command in Digital Signal Processing (DSP) which was also a best choice in noisy environment. The vocal command for Handicapped people wheelchair (HPWC) can be chosen. The voice command is used for guiding the wheelchair for impaired people suffering from spasms, paralysis etc. The joysticks cannot be used by paralyzed people and so application of joysticks failure such cases. The microcontroller can recognize the voice commands which were already stored in the database of AVG. For security concerns, four ultrasonic sensors were used to avoid the obstacles. The microcontroller can gives information to PC which in turn recognizes the voice command and makes decision for the movement of wheelchair. The preprocessing block is used to recognize the input signal by adopting it to the recognition system and also to minimize environmental noise along with the external noise originated from the wheelchair engine. The sound of the engine must be recorded and suppressed from the aggregate signal which includes the voice command combined with the sound of the motor in order to produce a signal to noise ratio which obviously increases the benefits and efficiency of the wheelchair. The speech location and hamming block were used to find out the start and end point of the words pronounced by the user. After that, the word is partitioned into a number of segments by eliminating silence. The output would be in the form of vector of word samples.

The speech location procedure is completely based on the analysing the zero point crossing and energy of the signal. The determination of the beginning and end of the words was done by linear prediction mean square error as it was computationally simple [9]. Francesco Beritelli *et al*. have proposed the algorithm VAD (Voice Activity Detection) which is used to support effective speech recognition on noisy environment. Nowadays the speech compression techniques are used in various fields like TV channels, FM, telephone services etc. In TV channels, the digital cellular system is used to increase the efficiency of speech coding system. Telephone service system can communicate through wired or wireless mode in which the background noise decreases the efficiency. Using VAD algorithm, the efficiency can be increased to a higher rate irrespective of the presence of external noise. Two types of algorithms were used to achieve this scenario, which are

i)   VAD standardized by ETSI for the GSM system and
ii) The VAD standardized by ITU-T for the 8-kbit/s G.729 speech coder.

Adaptive threshold mechanism plays a major role in VAD based on GSM system to make it act as an energy detector. Adaptive analysis filter was used for filtering the input signals and also to reduce the background noise. A new VAD was standardized by ITU-T in Recommendation G.729 annex B [10] [11]. It allows the ITU-T G.729 8-kbit/s speech coder to switch between two operative coding modes. If the VAD flag is one, the speech coder is invoked to code/decode active voice frames. If the VAD output is found to be zero, DTX/CNG algorithms are used to code/decode non-active voice frames [12].

3. Recognition of Partial Speech in Wheelchair

We propose the use of NAM for recognition of speech in wheelchair for impaired people. NAM microphone is a device used to recognize the quite murmur and the device can be attached behind the speakers' ear. Silicon is a thin layer that is in touch with the skin of the speaker to conduct NAM speech without external noise. Without air space, the silicon gets in touch with the talkers' mastoid which increases efficiency even in noisy environments. NAM is used to guide the wheelchair movements for people who can speak partial words with the help of an embedded microcontroller. The quite murmur is uttered by the speaker and can be detected using NAM microphone which cannot be heard by people nearby. The NAM microphone is robust for enabling personal conversation over mobile phones. The Tamil speech recognition engine (TSRE) can be used to recognize the words spoken by speakers using NAM speech stored in the TSRE database.

3.1. Partial Speech Recognize by NAM

People who have impaired vocal cords by nature or who have met any accident cannot speak full words. We proposed the NAM speech to be used for such handicapped people to command the wheelchair for its movement and NAM Speech is a soft murmur which is more efficient than the normal speech in the noisy environment even for people who can speak partial words. The waveform of partial speech recorded using NAM microphone is shown in **Figure 3**.

3.2. Rotation of Motor in Wheelchair

The motor plays a vital role in the movement and rotation of wheelchair. The microcontroller receives command from the user in the form of speech and the speech command is compared with the predefined commands stored in the TSRE dataset and the respective action is taken in accordance to that command. This leads to the wheelchair movement. The architecture of the framework based on NAM microphone is shown in **Figure 4**.

There are different methodologies for changing the state of the microcontroller. Considering the first bit of
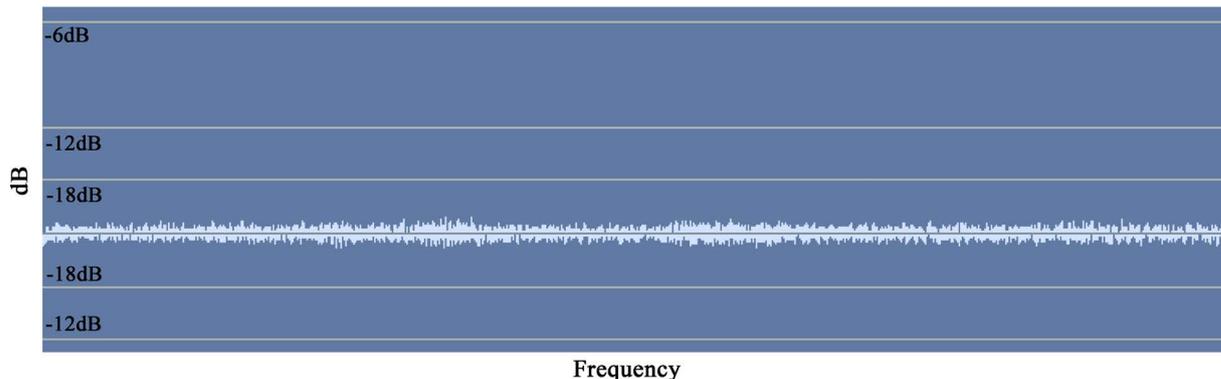

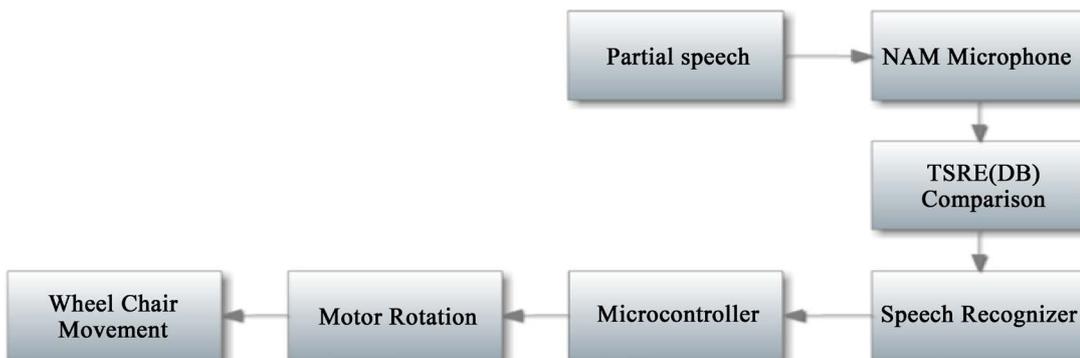**Figure 3.** NAM waveform of partial speech.

Figure 4. Framework using NAM speech for recognizing partial speech.

Port 1 to have the switch which controls the movement of wheelchair (Start command), the respective action is to be done on the first bit of Port 2. In order to achieve this, we can use Embedded C language to design commands which acts as an interface between the microcontroller and the motor to perform the required action. For the purpose of switching ON the switch, we have to OR the value 0X10 with the Port 1 value whereas if we are in need of switching it off, we can AND the value EF with the value available in Port 1. This scenario is shown in **Figure 5** and **Figure 6.**

We have used Arduino board microcontroller which is intended to make it easier to build an interactive objective environment. The microcontroller board consists of 6 analog input pins at the bottom right corner. In addition to that, there are 6 digital pins at the top right corner. The Arduino board is shown in **Figure 7**.

4. Experimental Results

The experiment is done in a noisy environment since NAM microphone is robust in noisy environment. The application of NAM speech can be used in voice controlled wheelchairs. Since there is no air space between the NAM microphone and the human mastoid, the clearance of each word is found to be robust. NAM microphone
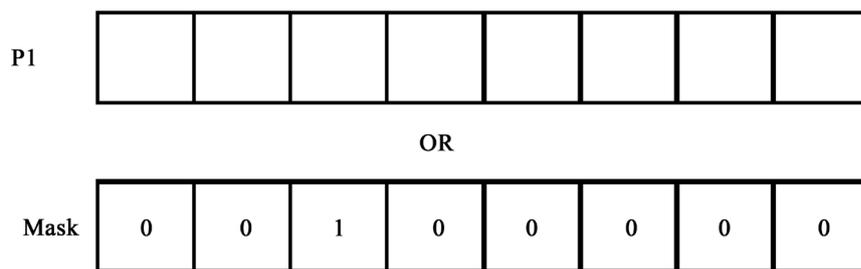
P1								

OR

Mask	0	0	1	0	0	0	0	0

Figure 5. Masking 0X10 for switching ON the fourth bit in port 1.

P1								

OR

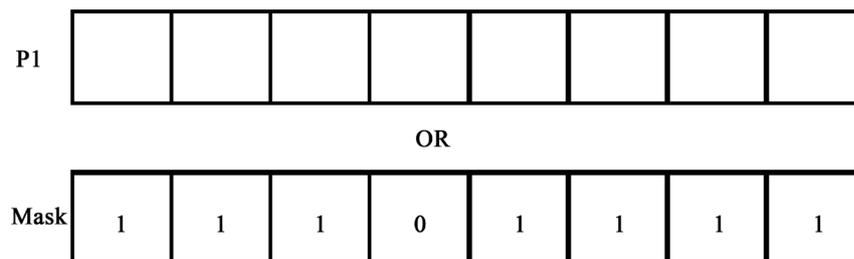| Mask	1	1	1	0	1	1	1	1

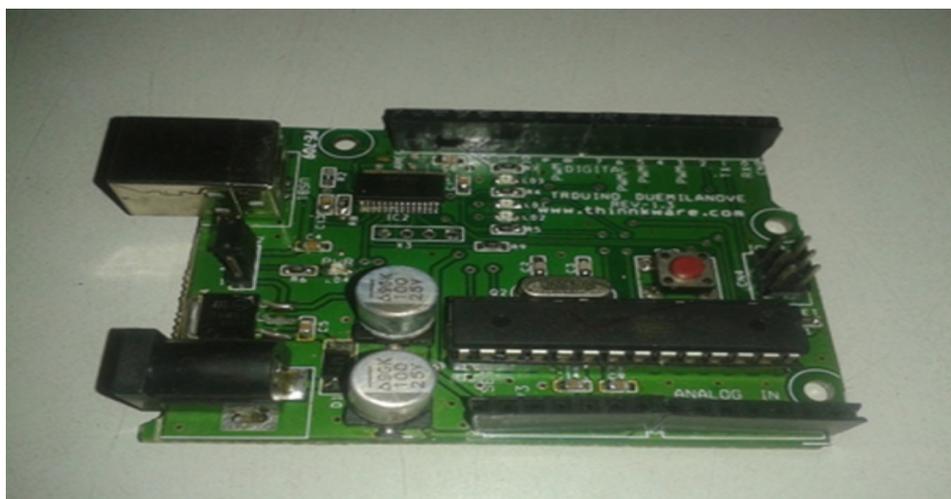**Figure 6.** Masking EF for switching OFF the fourth bit in port 1.

Figure 7. Arduino Microcontroller.

especially helps out people capable of speaking partially impaired by birth or affected by paralysis or any other kind of speech related disease. The microcontroller can recognize the voice command based on the code designed using KEIL interface using embedded C programming and then the microcontroller verify the speakers' murmur with respect to the words stored in TSRE dataset. The partial speech recorded using NAM microphone has been applied in voice controlled wheelchairs even for people who can speak partially and a compromising level of accuracy is reached. The overall connection done with the help of the Arduino kit along with the tachometer (used for controlling the rotation speed of the wheel by controlling the input voltage) and sound sensors is shown in **Figure 8**.

The assembled moving wheel with the motors representing the wheelchair which works on voice commands is shown in **Figure 9**.

The words used for guiding the wheelchair are shown in the **Table 1**. All the words available in the training dataset are recognized correctly and the words which are not in the training dataset are not detected and have no effect in the movement of wheelchair.

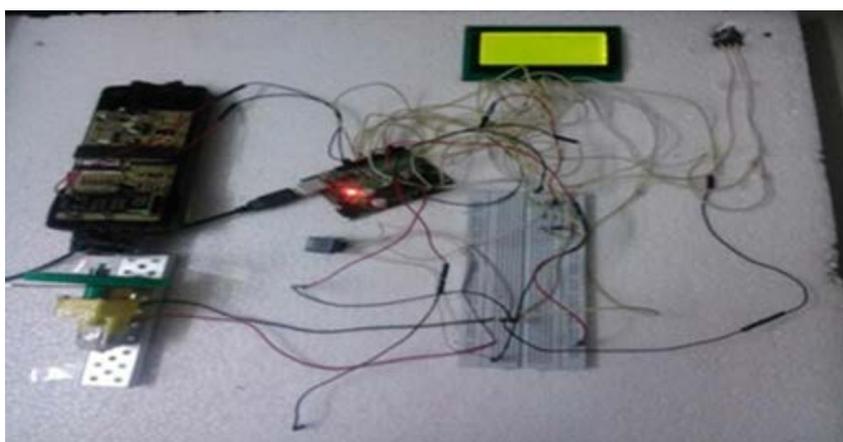**Figure 10** shows the comparison between the normal microphone and NAM microphone. It is understood

Figure 8. Overall connection using Arduino microcontroller and tachometer.

Figure 9. Moving wheel representing the wheelchair.

Table 1. Words used for wheelchair guidance.

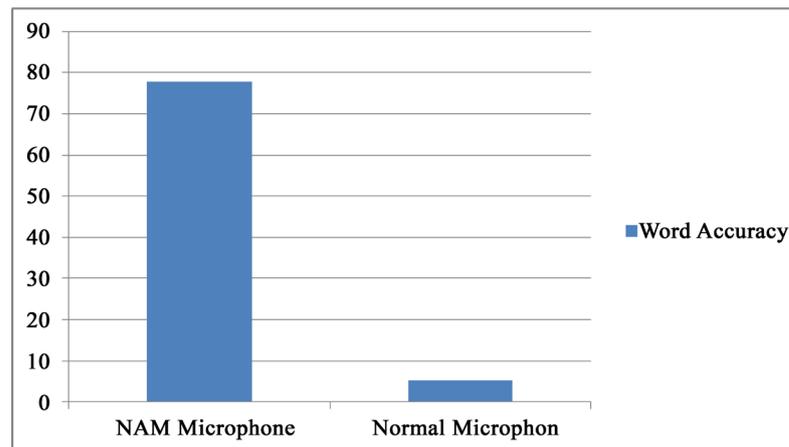| SI.NO	Sentence in DB	Words used for training in DB	Compare the DB with speaker voice	MAX correlated value	Wheelchair moves
1	முன்[FRONT]	முன்[FRONT]	yes	8.2529920	yes
2	பின்[BACK]	பின்[BACK]	yes	6.2932462	yes
3	நில்[STOP]	நில்[STOP]	yes	8.2688334	yes
4	வேகப்படுத்து [INCREASE SPEED]	வேகப்படுத்து [INCREASE SPEED]	yes	0.5248612	No

Figure 10. NAM and Normal microphone comparison.

from the figure that NAM microphone can produce better result compared to normal microphone. The accuracy level of normal microphone is less than 10% while NAM microphone can predict 70% of the word accurately.

5. Conclusion

In this paper, we have proposed and used NAM microphone for recognizing partial speech in paralysed people and integrated that into a wheelchair guidance system using TSRE (Tamil Speech Recognition Engine). This helps people who are paralysed and lost their voice, to move around in this wheelchair without the third person's help. The proposed method was designed and implemented in a wheelchair like model using Arduino microcontroller kit. Experimental results have shown that 80% accuracy can be obtained using this method, which is much higher than the normal microphone model. Though NAM microphone is robust to external noise, some of the noise cannot be cancelled. It is due to external noise factors like this the accuracy level was within 80%. Our future research work focuses on increasing the accuracy by cancelling the external noise factors and also integrating NAM microphone with normal microphone for recognizing. Further research is carried on in integrating lip movement with NAM microphone in order to improve the efficiency level of wheelchair guidance.

Acknowledgements

References

[1] Babani, D., Toda, T., Saruwatari, H. and Shikano, K. (2011) Acoustic Model Training for Non-Audible Murmur Recognition Using Transformed Normal Speech Data. 2011 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Prague, 22-27 May 2011, 5224-5227.

[2] Nakajima, Y., Kashioka, H., Shikano, K. and Campbell, N. (2003) Non Audible Murmur Recognition Input Interface Using Stethascopic Microphone Attached to the Skin. *Proceedings of IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing*, Vol. 5, 708-711.

[3] Kumar, A., ingh, P., Kumar, A. and Pawar, S.K. (2014) Speech Recognition Based Wheelchair Using Device Switching. *International Journal of Emerging Technology and Advanced Engineering*, **4**, 391-393.

[4] Ye, H. and Young, S. (2006) Quality-Enhanced Voice Morphing Using Maximum Likelihood Transformations. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **14**, 1301-1312.

[5] Mokbel, C. (2001) Online Adaptation of HMMs to Real-Life Conditions: A Unified Framework. *IEEE Transactions on Speech and Audio Processing*, **9**, 342-357. http://dx.doi.org/10.1109/89.917680

[6] Afify, M., Cui, X.D. and Gao, Y.Q. (2009) Stereo-Based Stochastic Mapping for Robust Speech Recognition. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **17**, 1325-1334. http://dx.doi.org/10.1109/tasl.2009.2018017

[7]   Cooper, R.A., Jones, D.K., Fitzgerald, S., Boninger, M.L. and Albright, S.J. (2000) Analysis of Position and Isometric Joysticks for Powered Wheelchair Driving. *IEEE Transactions on Biomedical Engineering*, **47**, 902-910. http://dx.doi.org/10.1109/10.846684

[8]   Zeng, Q., Teo, C.L., Rebsamen, B. and Burdet, E. (2008) A Collaborative Wheelchair System. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **16**, 161-170.

[9]   Fezari, M. (2007) Voice and Sensors Fusion for Guiding a Wheelchair. *Journal of Engineering and Applied Sciences*, **2**, 30-35.

[10] Benyassine, A., Shlomot, E., Su, H.Y., Massaloux, D., Lamblin, C. and Petit, J.P. (1997) ITU-T Recommendation G.729 Annex B: A Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications. *IEEE Communications Magazine*, **35**, 64-73. http://dx.doi.org/10.1109/35.620527

[11] ITU (1996) A Silence Compression Scheme for G.729 Optimized for Terminals Conforming to Recommendation V.70. ITU-T G.729 Annex B. International Telecommunication Union, Geneva.

[12] Beritelli, F., Casale, S. and Cavallaero, A. (1998) A Robust Voice Activity Detector for Wireless Communications Using Soft Computing. *IEEE Journal on Selected Areas in Communications*, **16**, 1818-1829. http://dx.doi.org/10.1109/49.737650