Human Computer Interaction System for Impaired People by using Kinect Motion Sensor : Voice and Gesture Integrated Smart Home

K. A. S. V. Rathnayake[1], W. K. I. L. Wanniarachchi[2*]
Department of Physics
University of Sri Jayewardenepura
Nugegoda, Sri Lanka
s.vimukthini@gmail.com[1], iwanni@sjp.ac.lk[2]

W. H. K. P. Nanayakkara[3]
CodeGen International (Pvt) Ltd.
Colombo 10, Sri Lanka
[3]hansa.nanayakkara@gmail.com

***Abstract*— Gesture and speech based Human computer interaction is one of the most natural and convenient way of communication which leads to minimize the gap between human and machine. In this paper, we propose a single HCI system which can be utilized by people with physical challenges as well as speaking and hearing disabilities. Voice and gesture commands acquired by Kinect motion sensor are availed to manipulate home appliances via a Wi-Fi enabled wireless network hub. Even though nowadays, there is a pool of HCI systems with various technologies, cost and accuracy of those systems are debatable factors. Experiments carried out in the evaluation process reveal that our system provides more than 80% of recognition accuracy in gesture mode as well as in voice mode. Moreover, it can work well in uncontrolled, environments.**

Keywords—Microsoft Kinect V2; HCI; Gesture recognition; Voice recognition; Smart home; NodeMcu.

I. Introduction

Developing Human-Computer Interaction (HCI) System means assuring usability and functionality of the system by giving tremendous support for users in interaction as well as providing the amiable user experience. In recent few years many researchers have been focusing on this field to emerge with novel technologies to enhance the human computer interaction in more natural way. In order to obtain a good HCI system it needs to have several goals to accomplish such as; creating an efficient and safe interaction system, it should be easy and enjoyable to use while putting people first. If it cannot full fill these factors, it will no longer provide a good solution [1]. The proposed HCI solution mainly addresses disabled people who need assistance in their day to day life.

According to World Health Organization (WHO) estimates that, from whole world's population near 15% of the people spending their life with some kind of disability and among them, 2% to 4% of people are with noticeable difficulties in functioning. Results of World Health Survey also reveal that disability prevalence is higher in developing countries than in developed countries. However, estimates for various disabilities are progressively rising day by day due to many reasons such as; crucial problem of population aging, chronic diseases which are being steeply permeate and development of measuring disabilities [2].

Many of their life barriers are avoidable and the difficulties they face can be defeated. The goal of this proposed work is to provide support to break some barriers of their life and to assist them with their daily life activities. We introduce this HCI system with high usability, high accuracy and low cost. HCI system works for user gesture commands and voice commands by using Microsoft Kinect V2 which is used as the motion sensor and the audio sensor. The commands received by NodeMcu through voice and gesture recognition section of the system are used to control appliances in the home or any other place wirelessly. This open source system is appropriate for people with hearing disability, speaking disabilities as well as for people who have the only ability of speaking.

II. Related Work

HCI systems have been developed by using various detection methods, recognition methods as well as communication methods in order to overcome its challenging problems. Existing systems can be generally split into two groups such as vision-based approaches and non-vision-based approaches according to its detection methods.

Vision based approaches face problems of quality of image received from optical cameras as well as the variation of distance to the sensor [3], [4]. In their work, input gesture images were captured via a web camera or any camera and further processed in MATLAB environment to extract the hand from the background. Proportion of researches using non-vision-based approaches than vision-based approaches is higher, as system highly depends on lighting condition and there are good as well as ample of sensors and devices in the market to conduct non-vision-based approaches.

Under different lighting as well as noisy conditions Ultrasonic sensing technology tracks dynamic gestures and it measures the distance between point and the sensor by using ultrasonic signal that reflects from the point. When the gesture is being performed, it takes the coordinates of the location of the hand. Nevertheless, the accuracy of ultrasonic sensor can be affected by humidity and temperature of surrounding [5]. Liu et al. used four such sensors mounted on four sides of an android tablet to gather data coming from four sides. As data come from four different sides simultaneously, it can create six degrees of freedom gesture in 3D format. Their system is able to identify 12 different gestures. Similarly, variation of Radio frequency and electric field are employed to detect hand movements where Kellogg et al. used Radio-frequency identification tags (RFID) to track hand movements [6] and Kim and Moon utilized an Electric Potential Integrated Circuit (EPIC), an AC coupled device to obtain variations of electric field with body movements [7]. EPIC can be interfered by other external electric field and also it cannot be used to create strong fields either, therefore system limits for shorter distances. Non-vision-based systems are also appeared with wearable devices. WiiRemote which has in built accelerometer sensor is such device many researchers use in their gesture recognition systems to collect time series data of acceleration of each movement. [8], [9], [10], [11].

Even though HCI systems are being developing for voice recognition and gesture recognition, the critical problems that systems facing are, accuracy of system recognition and system feasibility with disabled people. HM2007 IC as a trained voice recognition module was availed by P. Gupta [12] and commands were stored in binary format, meanwhile those commands were sent through 8-bit data bus to Atmega16 microcontroller to perform specific task. Rather than using ICs for recognition, recognition algorithms have been entering to the field to enhance the accuracy of recognition results as well as to improve the system feasibility. Research work have been carried out in references [13], [14], [15] are several systems based on MATLAB algorithm where some were aimed for automating home appliances [13], [14]. Hidden Markov Model (HMM), Support Vector Machine (SVM), Dynamic Time Wrapping (DTW) are rampantly using algorithm for both gesture and voice recognition systems [7], [9], [10] because of their high performance. Meanwhile, scanning method algorithm and other different customized classification methods were also used with gesture recognition systems [3], [4].

Method of communication with devices also provides HCI system a significant contribution to make the whole system worthwhile. Various communication options have been being experimented for controlling appliances such as mobile network, Bluetooth, RF, Wi-Fi, Internet, LAN technologies and so on. [16] In a place where Wi-Fi is not available, ZigBee, RF module was employed to communicate with appliances via remote controller. However, nowadays Wi-Fi is available almost everywhere, in that context; this system will not get the benefit of available resources. System based on Bluetooth communication method [18] gathers data from hand phone which hosts a Python script. It sets up an ad-hoc to communicate between devices and Arduino BT board. Even though single shared Bluetooth module reduces the wiring cost, it leads to several other disadvantages such as incurring access delay, low data rate as well as shorter data range which makes the system less feasible. Moreover, systems based on Wi-Fi [19], [20] provide high data rate and long data range. Nevertheless, it requires a mobile phone as an external device in the system.

In above, we have addressed and critically reviewed many difficulties and issues of past works related to our proposed work. Compared to established systems, this proposed prototype is low cost, highly accurate and user friendly which are considerable factors specially for physically impaired people. The system basic needs are WiFi connectivity as well as having Microsoft Kinect Sensor which is accountable for extracting high quality vocal and gestural inputs. Setting-up the module is effortless for users as it does not require rewiring and facilitates to establish with any plugging location. In addition, IR sensor is availed by Kinect sensor to track skeleton movements. Therefore, system accuracy is independent of lighting condition. Furthermore, since we are focusing on disable people, a system that doesn’t require any physical touch and having both facilities of speech recognizing as well as gesture recognizing is an optimal solution.

III. Microsoft Kinect 2.0

Kinect is a device for human–machine interaction, which adds two more input modalities to the palette of the user interface designer: gestures and speech [21]. Kinect is one of the major parts of the proposed system. It consists of four microphone array which captures audio data at a 24-bit resolution, at 16 kHz [22]. Its audio capabilities such that echo cancellation, beam formation and sound localization cause the high resolution and support to boost the system recognition accuracy. Moreover, IR emitter and a monochrome CMOS (Complementary Metal-Oxide Semiconductor) sensor work together to see objects in 3-D regardless of the lighting conditions and captures VGA resolution video as well as depth images [23]. Its vertical field of view and horizontal field of view are 60^0 and 70^0 respectively and it has depth range between 0.5m to 4.5m.

IV. Implimentation

A. Hand Detection and Gesture Recognition

Skeleton of the body is detected by Kinect when it sees the head and upper body of the user; no need of particular posture or calibration to be taken by the user. The IR emitter projects pattern of IR light and this pattern of light is utilized for depth calculation of the user in the viewing range of the sensor to identify user’s body parts [24]. The database of gesture commands was trained with the aid of Microsoft Kinect Studio and Visual Gesture Builder (VGB). Kinect studio was mainly used to recode hand movement for each defined gesture command and those were fed into VGB where gesture database is been created using machine learning algorithm. Detection approach of the proposed system is discrete detection and VGB uses AdaBoostTriger detection technology in Adaptive Boosting (AdaBoost) machine learning algorithm to decide whether user is performing the gesture or user is not. Hand gestures are recorded and each recorded video clip is tagged with a boolean value which represents the gesture occurrence.

B. *Voice Detection and Speech Recognition*

Audio stream data coming through microphone array are subjected to Kinect audio prepossessing in order to achieve quality vocal input. The speech model which was developed for the system utilized Microsoft Speech Grammar and Speech Runtime to recognize the vocal input. Speech Grammar consists of a set of grammar rules and it helps Speech Recognition Engine in Speech Runtime to recognize the vocal input meaningfully.

The speech model as well as gesture model provides all the information based on confidence value that portrays the certainty of the result. It depicts how good the speech command and gesture command has been identified by Kinect. A threshold value for confidence is set as per need; considering the required recognition accuracy. Thereafter compares its value with the confidence value generated by the model at each performance. If the confidence value is higher than the threshold value, the event handler of the model will be appointed and GUI will be manipulated according to the received output.

C. *System Framework*

Fig. 1 depicts overview of the system which mainly uses two modules; speech recognition and gesture recognition. Audio stream data as well as gestural data coming from Kinect sensor are subjected to process in preliminary processor in order to obtain high quality input data. Processed data are carried to develop both speech model and gesture model. Output of speech model and gesture model are communicated to appliances through NodeMcu which is housed by remote module.

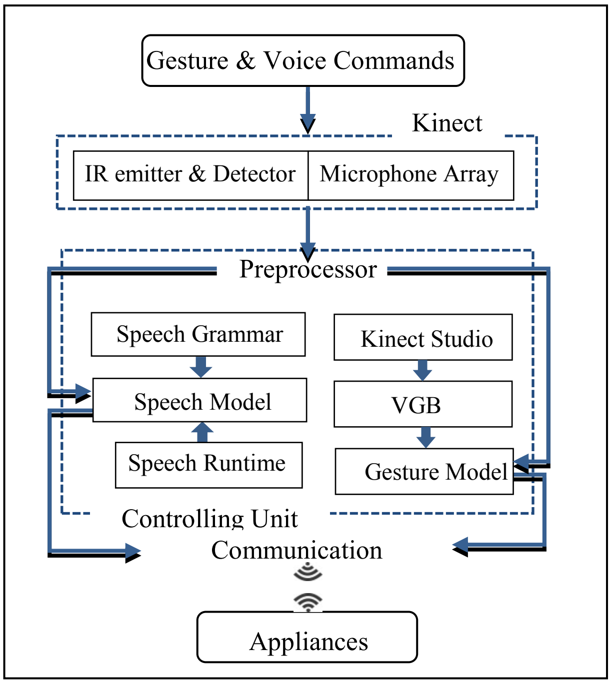

Fig. 1. Network topology of the system

D. *Network Architecture*

Fig. 2 illustrates the network topology of the system and connection between controlling unit and the other components of the system.

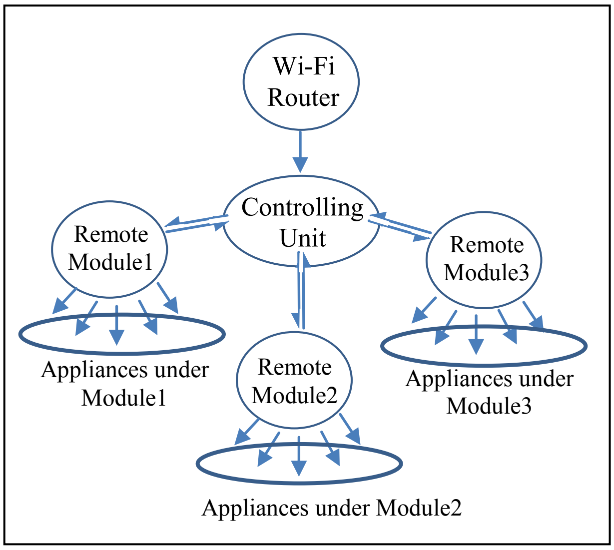

Fig. 2. Network topology of the system

The system can be configured by replacing existing switch with remote module. Thus, it facilitates to install the system without changing internal wiring of existing house.

Wi-Fi enabled controlling unit is the heart of the system which coordinates each remote module according to the command received from the user. Remote module consists of a NodeMcu and an Opto-isolator relay module where each NodeMcu acts as a server as well as a client. Controlling unit that runs a server communicates with non-smart appliances through the relevant remote module and smart appliances are manipulated directly through controlling unit.

E. *NodeMcu*

NodeMcu is a ready-made open source development board with ESP8266-12E Wi-Fi module which works under eLua based firmware. It is embedded with Tensilica L106 32-bit microcontroller unit and 16-bit RSIC which is reaching a maximum clock speed of 160 MHz [25]. This high performance, low cost, ultra-low power consumption wireless System-On-Chip can use network protocols; IPV4, HTTP, FTP, UDP & TCP.

V. System Graphical User Interface

The difference in our GUI is that it enables to control a lot of appliances in many rooms by keeping its user friendliness in a high level as our target group is disabled people. Voice mode works as the default command mode and whenever you need to shift the mode from voice mode to gesture mode and vice versa GUI can be used to perform it. Initially administrator is required to register with the system by simple adding his or her details, names of appliances, number of rooms and their names. Other users can be registered with the system by following the same through the administrator. Since the design of the GUI is simple and easy to understand, anyone can follow that providing steps to manipulate registered appliances in the home.

VI. Gesture Comands

Fig. 3, shows the gestures used for controlling system GUI as well as appliances; On, Go Back, Selecting, Off and Scrolling down.

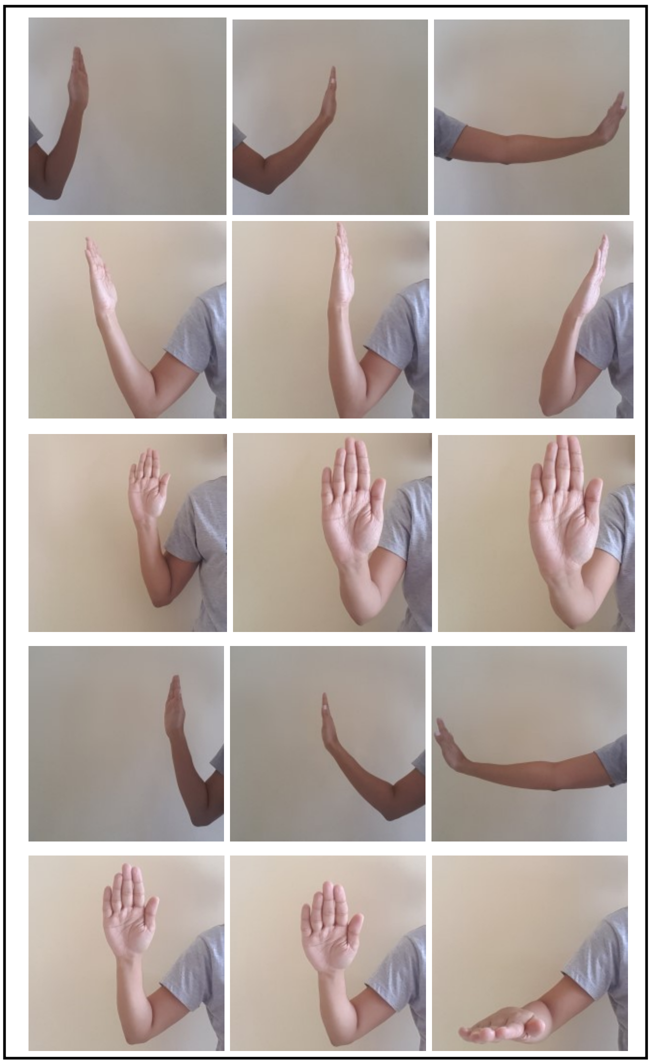

Fig. 3. Gesture commands

VII. Performance Evaluation

To test the performance of the proposed HCI system, voice mode and gesture mode of the system was evaluated separately. In our previous work we proposed voice controlling system which can be used to govern appliances in a home [18].

The experiment was initialized in an isolated room of dimension 6 m × 6 m × 12 m. As input commands, recorded commands were used at sound level of 100.0 dB ± 0.5 dB as it leads to reduce human errors. A uniform audio sound which was placed at 3 m away from Kinect was varied into three levels and they were measured with respect to the sensor. At each that external noise level, performance distance for each command and accuracy of Kinect for that each performance distance were obtained. The performance distance was varied from 1 m to 4 m keeping sound level of the command and noise level constant. The program was restarted for each distance as it adapts to the user's speech pattern. Experiment was repeated twice in order to increase the accuracy and repeated again by decreasing noise level from high to low. The experiment was executed in background noise level 47 dB ± 1 dB. Following figure depicts how recognition accuracy varies with performance distance. Level I, Level II and Level III are noise levels which were measured during the experiment with respect to the Kinect sensor.

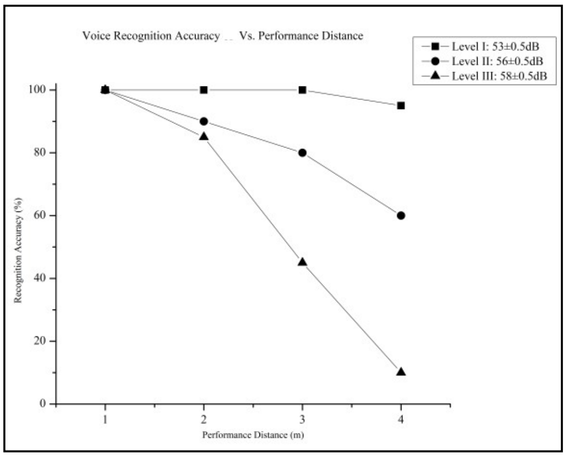

Fig. 4. Graph of voice recognition accuracy vs.performance distance

As shown in the line graph (Fig. 4), recognition accuracy gradually drops with existence of a high noise level. But it is a noticeable feature that even when the performance distance and source of external noise are at the same distance from Kinect, the system still gives 45% accuracy.

To evaluate the performance of gesture mode, an experiment was performed in a room where luminous intensity 50 lux – 53 lux. It was proceed with three people and they performed five gesture commands contained in the database, ten times at three different speeds varying the performance distance 1 m to 4 m. Speed levels were classified as slow, medium and high according to time duration 0.5 s, 1.0 s and 2.0 s respectively.

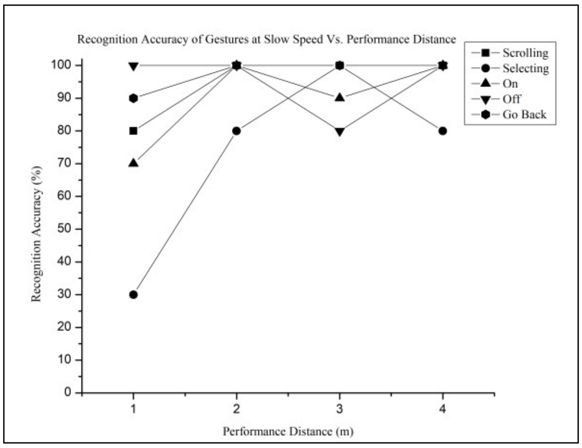

Fig. 5. Graphs of recognition accuracy at slow speed vs. distance

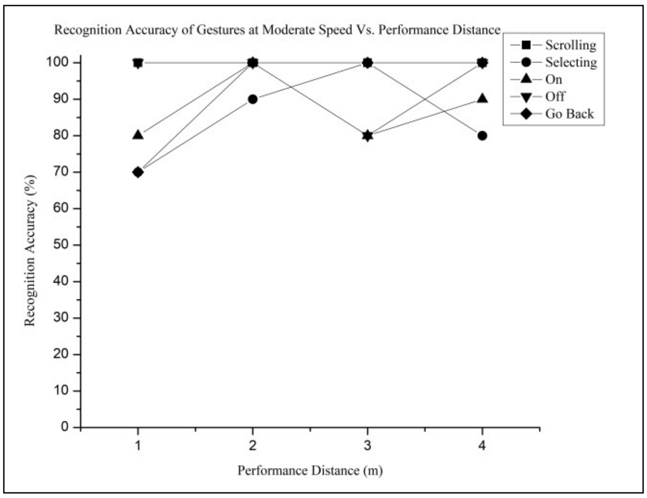

Fig. 6. Graphs of recognition accuracy at moderate speed vs. distance

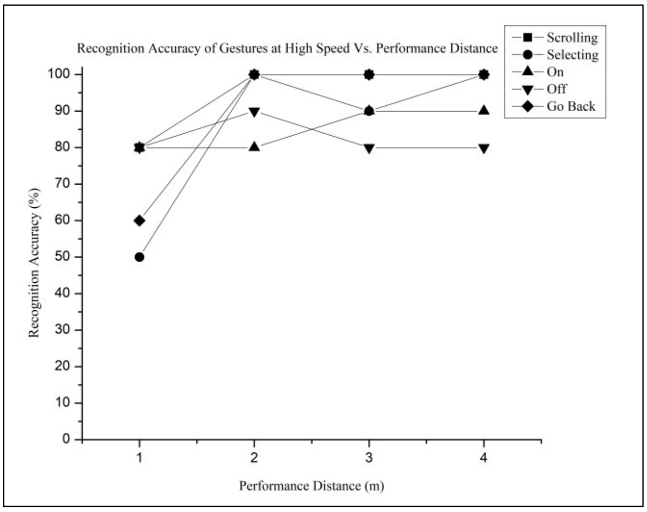

Fig. 7. Graphs of gesture recognition accuracy at high speed vs. distance .

According to Fig. 5, Fig. 6 and Fig. 7 recognition accuracy at every speed upsurge until performance distances about 1 m and maintains more than 80% accuracy until about 3.5 m. When "Selecting" gesture is performed at 1 m distance, gesture crosses Kinect vision range which starts from 0.5m and disrupts output accuracy. Thus "Selecting" gesture shows more than 80% accuracy between 1.5 m to 4 m at each speed and all the other gestures display more than 80% accuracy between 1m to 4 m at each speed.

VIII. Conclusion

HCI is prominent and rapidly developing research area and researchers draw their attention more on reducing the gap between Human and Machine. In that context, voice recognition and gesture recognition acquire special consideration of researchers as it is the most natural and convenient way of communicating and interacting with hardware systems. Our focus in this proposed work is by using one integrated system, addressing and providing convenience for both groups of people who are suffering from hearing, speaking disabilities as well as physical disabilities.

Fig. 8. Implementation of proposed system.

According to the performance evaluation it is apparently seen that, while voice recognition accuracy gives more than 80% accuracy from 1 m to 3 m even at high noisy back ground, gesture recognition provides more than 80% accuracy from 1.5 m to 3 m for all the gestures with all speed levels. In order to boom the recognition accuracy furthermore, system need to be enriched with huge data base.

For further study, face recognition will be integrated with the system, as it can raise the security level of the existing system. Moreover, to increase the convenience level of the user, system is supposed to be customized with user commands and also it leads to widen the system furthermore.

Acknowledgment

The authors would like to thank the University of Sri Jayewardenepura (Grant No ASP/01/RE/SCI/2016/25) for the funding support.

References

[1] "Introduction to HCI", *School of Computer Science, University of Birmingham*, 2018. [Online]. Available: https://www.cs.bham.ac.uk/~rxb/Teaching/HCI%20II/intro.html. [Accessed: 04- Jan- 2018].

[2] "WHO | World report on disability", *World Health Organization*, 2018. [Online]. Available: http://www.who.int/disabilities/world_report/2011/report/en/. [Accessed: 09- Feb- 2017].

[3] P. Chowdary, M. Babu, T. Subbareddy, B. Madhava and V. Elamaran, "Image Processing Algorithms for Gesture Recognition using MATLAB", in *International Conference on Advanced Communication Control and Computing Technologies*, 2014, pp. 1511-1514.

[4] M. Panwar, "Hand gesture recognition based on shape parameters," *2012 International Conference on Computing, Communication and Applications*, Dindigul, Tamilnadu, 2012, pp. 1-6.

[5] Y. Liu et al.(2013), SoundSense: 3D Gesture Sensing Ultrasound on Mobile Devices [Online].Available: http://mrorz.github.io/files/soundsense.pdf

[6] J. Smith, T. White, C. Dodge, J. Paradiso, N. Gershenfeld and D. Allport, "Electric field sensing for graphical interfaces," in *IEEE Computer Graphics and Applications*, vol. 18, no. 3, pp. 54-60, May/Jun 1998.

[7] Y. Kim and C. Moon, "Non-Contact Gesture Recognition Using the Electric Field Disturbance for Smart Device Application", in *International Journal of Multimedia and Ubiquitous Engineering*, Vol 9, Issue 2, 2014.

[8] J. Wu *et al.*, "Gesture Recognition with a 3-D Accelerometer", in *Proc. 6th on Ubiquitous Intelligence and Computing Int. Conf.*, Brisbane, Australia, 2009, pp. 25-38.

[9] A. Akl, C. Feng and S. Valaee, "A Novel Accelerometer-Based Gesture Recognition System," in *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6197-6205, Dec. 2011.

[10] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *2009 IEEE International Conference on Pervasive Computing and Communications*, Galveston, TX, 2009, pp. 1-9.

[11] H. Guesgen and D. Kessell, "Gestural Control of Household Appliances for the Physically Impaired", in *International Florida Artificial Intelligence Research Society Conference*, 2012, p. 353.

[12] P. Gupta, "Human Voice Controlled Home Appliances", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 5, pp. 142-143, 2015.

[13] S. Amrutha, S. Aravind, A. Mathew, S. Sugathan, R. Rajasree and S. Priyalakshmi, "Voice Controlled Smart Home", *International Journal of Emerging Technology and Advanced Engineering*, vol. 5, no. 1, pp. 272-275, 2015.

[14] M. Katore and M. R. Bachute, "Speech based human machine interaction system for home automation," *2015 IEEE Bombay Section Symposium (IBSS)*, Mumbai, 2015, pp. 1-6.

[15] A. Hussein, "Analysis of Voice Recognition Algorithms using MATLAB", *International Journal of Engineering Research & Technology*, vol. 4, no. 8, pp. 273-278, 2015.

[16] K. Gill, S. H. Yang, F. Yao and X. Lu, "A zigbee-based home automation system," in *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 422-430, May 2009.

[17] K. A. S. V. Rathnayake, S. I. A. P. Diddeniya, W. K. I. L. Wanniarachchi, W. H. K. P. Nanayakkara and H. N. Gunasinghe, "Voice operated home automation system based on Kinect sensor," 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), Galle, 2016, pp. 1-5.

[18] S. I. A. P. Diddeniya, K. A. S. V. Rathnayake, H. N. Gunasinghe, W. K. I. L. Wanniarachchi, "Integrating IRobot-Arduino-Kinect Systems: An Educational Voice and Keyboard Operated Wireless Mobile Robot " International Journal of Scientific Engineering and Technology, vol. 6, no. 6, 2017, p. 188., doi:10.5958/2277-1581.2017.00017.1.