39

Voice Operated Home Automation System Based on Kinect Sensor

K. A. S. V. Rathnayake, S. I. A. P. Diddeniya, and W. K. I. L Wanniarachchi
Department of Physics
University of Sri Jayewardenepura,
Nugegoda, Sri lanka.
iwanni@sjp.ac.lk

W. H. K. P. Nanayakkara
CodeGen International (Pvt) Ltd.,
Colombo 10, Sri Lanka.

H. N. Gunasinghe
Department of Mathematics and Computer Science
Open University of Sri Lanka
Nawala, Sri Lanka.

***Abstract*—Control of home appliances using smart technologies which is also known as home automation is a popular (industrial) research area. Many of these systems use remote controlling such as IR sensing, networking, Arduino programing etc. whereas our system uses voice controlling. Proposed system is designed to enable centralize controlling over distant household appliances. In this prototype, we built an interface demonstrating voice control using Kinect V2 as voice receiver and trained a computer system to identify set of voice commands. Then a circuit was constructed using Arduino and light bulbs that mimics actual appliances. We measured accuracy of our system. It is more than 95% when the distance between user and Kinect sensor is 4 m and when there is about 53 dB noise. Hence, the proposed method can be efficiently used with a Kinect V2 for voice controlling.**

Keywords—Microsoft Kinect V2; home automation; speech recognition; device control; voice control

I. INTRODUCTION

Speech recognition is a very familiar concept in today's world. This concept is used in many applications including Google Voice Search, iPhone's voice-operated assistant Siri, automated phone queues, in ticket-booking services, virtual assistant, Home automation system etc. [1]. Microphones and speech recognition software are used to build these applications.

Every person desires to spend a comfortable life. Therefore, a home automation system is very useful in the world now. Home automation system refers to the automatic and electronic control of household features, activity, and appliances. Some components of an automated home may include the centralized control of security locks on doors and gates, appliances, windows, lighting, surveillance cameras and HVAC systems (Heating, Ventilation and Air Conditioning). At present, there are many types of home automation systems in many countries already implemented and testing. Most of the existing systems face four main challenges; these are high cost of ownership, inflexibility, poor manageability, and difficulty achieving security. In this paper, we introduce a voice control home automation system with low cost, very flexible, high manageability, high efficiently and open source. In our system Microsoft Kinect sensor V2 is used as the only audio sensor and Microsoft Speech Recognition is used in speech recognition. Voice commands are converted to controlling signals using an Arduino program. The controlling signal of electrical and electronic devices is passed through an Arduino circuit. This system is particularly important to disable people to live a comfortable life. They can control electronics and electrical devices in home using their voice commands easily compared to using a remote control unit which they need to carry all time.

II. LITERATURE REVIEW

Existing, well-established systems are based on wired communication. Examples include BACnet, LonWorks and KNX [2], Neng-Shianget et al [3].Though these systems are ideal for large buildings they are very expensive to be used in a small household. Since the method proposed in this study can be applied at a lower cost, it is more suitable and effective for a small home or an indoor area.

In terms of wireless communications between devices [4] proposed an automation system based on Bluetooth without the internet controllability. Furthermore, some studies in [5, 6] contain examples for web based automation. However, they are not feasible to be implemented at a low cost solution. [7] K.G.H.A. Somasiri et al. [8] has implemented a low cost wireless web based home automation system locally. This home automation system is capable of turn on and off lights and different appliance connected to wall plugs, door locking and unlocking, shading level control and monitor the status above parts. Central control unit is connected to the internet all the time with the use of web server runs by Arduino Ethernet shield fixed in it. All the sub units are connected wirelessly to the main unit via RF transceivers and they communicate both way.

In another system P. Gupta [9] takes the input from voice recognition module and use microcontroller's intelligence to operate different devices. But in our system, we use Arduino

978-1-5090-6132-7/16/$31.00 ©2016 IEEE.

circuit to operate different devices. Therefore, we can build our system easily and more devices can be controlled.

S. Amrutha et al [10], used MATLAB to implement the voice controlled smart home system. MATLAB is not open source software. Therefore, this cannot be considered as a low cost solution. But in our system open source software (Visual C#, Visual C and Ardiuno IDE) and inexpensive equipments (Adiuno and 4 channel relay module) are used to implement the system.

In all above systems, they used normal microphones to capture audio data that are noisier. It directly affects the accuracy of these systems. But the developed system in this work, we use Microsoft Kinect V2 sensor that can capture audio data with reduced noises. Hence our system is reliable and accurate.

III. Implementation

In order to implement the Device Control Interface, Kinect version 2.0 is used with Visual C# programming. The configuration of the system is shown using the following diagram.

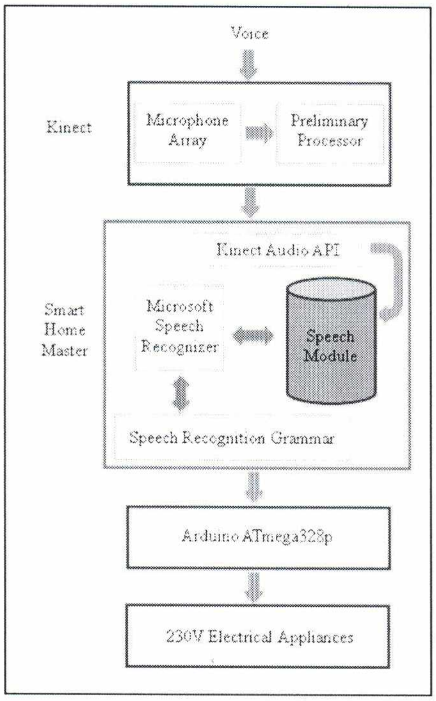

Fig. 1. Flowchart of the system

A. *Kinect Sensor Version 2.0*

Microsoft used this technology for playing games in the beginning. But at present the technology has been applied to real world applications. Kinect sensor consists of four main components as shown in the following figure 2. Among them, microphone array component is used in this proposed audio project.

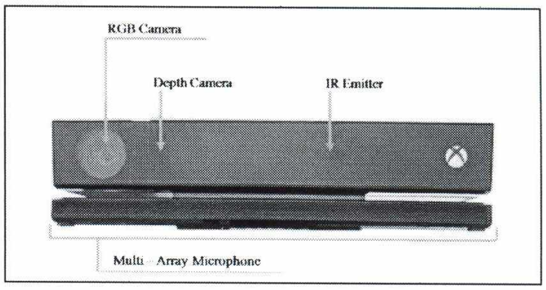

Fig. 2. Kinect Sensor

B. *Kinect Microphone Array*

The Kinect sensor has an array of four linear microphones and has 24-bit audio data resolution [11]. This microphone array uses Analogue to Digital Converter and specially provides in-built signal processing. It pre-processes audio which removes background noise from the captured voice signal and also removes echoes of the signal using automatic echo cancellation (AEC) algorithms in order to increase the quality of the audio data. Noise reduction and echo cancellation can be done effective with microphone array rather than a single microphone [12].

When there is a multiple set of microphones, the time that sound arrives from an audio source to each microphone is slightly different. Audio data captured from Kinect microphones go to the preliminary processor and at that section Beam forming and Sound localization techniques determine the direction of the sound source and the set of microphone is used as a directional microphone. Each sound coming from an appropriate source (Beam) is split in to approximately 24 frames per second. Then these inputs go to the PC for further processing.

C. *Central Controller (Smart Home Master)*

Central controller (Smart Home Master) is the heart of our system, Device Control Interface. Audio stream data coming from Kinect will be further processed in this section. Microsoft Kinect API will be used to construct the speech module which drives our audio system.

D. *Speech Module*

In the Speech module section, we have used Microsoft Speech Recognition Engine (Speech recognizer) and Microsoft Speech Recognition Grammar to implement the project.

E. *Microsoft Speech Recognition Grammar*

A speech recognition grammar contains one or more rules. Each rule defines a set of language constraints that a speech recognition engine uses to restrict the possible word or sentence choices during the speech recognition process [13].

F. *Microsoft Speech Recognition Engine*

Microsoft provides a speech recognition engine that matches vocal inputs with the words and phrases, defined by grammar rules of speech recognition. There are simple and complex rules designed. Simple rules recognize small set of words and phrases. A more complex grammar can recognize and organize semantic content from a variety of user accents. A speech recognition grammar also defines a set of grammar specific properties, such as locale, semantic format, and mode [13].

Microsoft speech recognition engine handles the part of speech recognition with a confidence value which lies between 0 and 1. A given confidence threshold value is compared with the confidence value which depicts the certainty of the selection. If that confidence value is greater than or equal to the given confidence threshold value, the command will be executed. We have set it to a low value in the system. The suitability of this value with our system will be tested in the experiments and results section.

G. *Arduino Uno*

The Arduino Uno is a microcontroller board based on the ATmega328p. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller. We either need to connect it to a computer using a USB cable or power it with an AC-to-DC adapter. The Arduino circuit acts as an interface between the software part and the hardware part of the project [14].

Speech module sends the recognized command to the Arduino Uno as a message to the serial port. That transmitted message is matched with the predefined commands and the selected command is executed. This system can be changed at any time by defining new commands, new appliances and by increasing number of appliances. In our proposed work we have used tentative four outputs which are Bedroom room bulb, Bedroom fan, Kitchen oven and Kitchen air fryer. The appliances are connected to the Arduino Uno pin numbers 2, 3, 4, and 5 respectively via the relay board. When the commands are received in to the Arduino Uno, relevant pins will be high or low according to the received command.

H. *Relays And Optoisolators*

Relays are used as electromagnetic switches that can turn a much larger current on and off. In the circuit of this system four relays are used to control four home appliances in home which needs high voltage 230V or high current to operate. Moreover, for further protection, optical isolators are used along with relays which are typically used to prevent flowing unwanted feedback into the low voltage circuit by keeping both sides electrically isolated from each other. Following figure 3 shows the implementation of electrical appliances with Arduino Uno and opto isolated relay module.

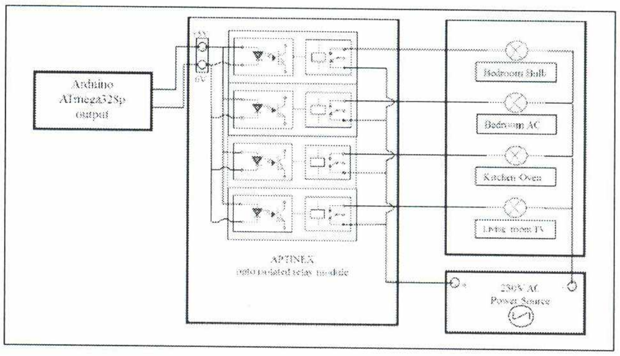

Fig. 3. System Implementation

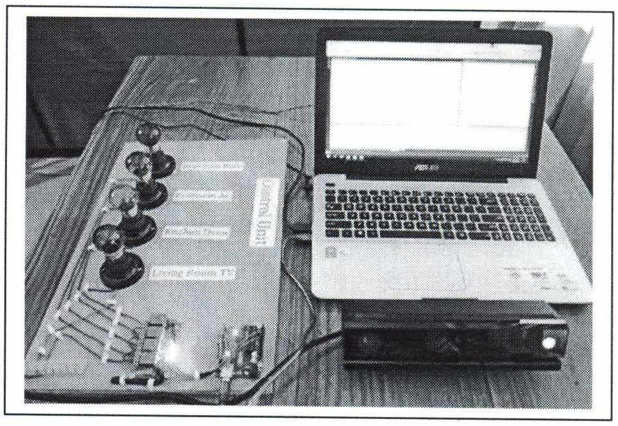

Fig. 4. Developed System

I. EXPERIMENT AND RESULTS

Device Control Interface was tested with Windows 10 64-bit operating system and used British Language pack. An experiment was conducted and data analysis was made to test the proposed work in order to get an idea of the system accuracy and how relevant it is to the practical world. Our tested environment was isolated and it was assumed as user bed room with dimension 6 m × 6 m × 12 m.

The experiment was implemented with using two parameters. They are performance distance at the specific noise level and the accuracy of the Kinect output for the command. The performance distance was taken by measuring the distance between Kinect and the place exactly where the recorder was placed. A recorded command was used for the experiment in order to reduce human errors and its sound level was 100.0 dB ± 0.5 dB. A constant audio sound was used as the experiment noise and it was placed 3 m away from the Kinect sensor. The experiment was conducted under the background noise range 47 dB ± 1 dB.

The performance distance was increased from 1m to 4 m keeping the recorded audio volume and the noise level constant. The experiment was repeated twice for each distance level to increase the accuracy of the data. The size of a data set was limited to 20 and the program was restarted for each data set as it adapts to the user's speech pattern. The same experiment was repeated by decreasing the noise level from

100 dB, 95 dB and 90 dB at the noise source. Following figure shows the variation of recognition rate with the performance distance.

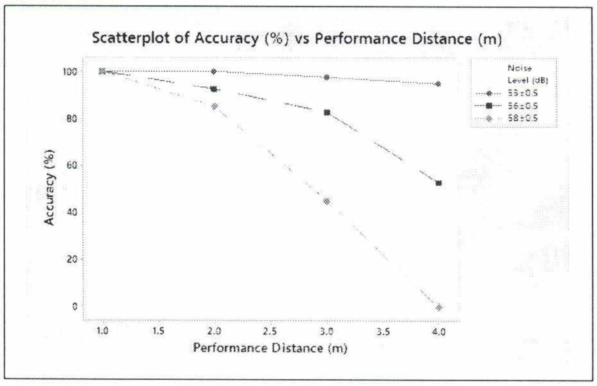

Fig. 5. The graph of accuracy vs. performance distance

In the above graph blue line indicates added noise level 53.0 dB ± 0.5 dB, red line indicates added noise level 56.0 dB ± 0.5 dB and green line indicates added noise level 58.0 dB ± 0.5 dB as sense by Kinect. The graph shows the accuracy of Kinect output for the command rapidly decreases with the performance distance.

The sound level of the noise as sense by Kinect and the sound level of the recorded command as sense by Kinect are same at the point 3m. Even when sound level of noise and sound level of command which are sensed by the Kinect are equal, the accuracy of that point is about 45%.

II. Discussion

British accent is more similar to Sri Lankan people's English accent. Therefore, British Language pack is used in the program. The speech recognition algorithm adapts to user's speech pattern and when it gets more data it adapts to the user better. If adaptation of the acoustic model is **on** updated value is 1 and if adaptation of the acoustic model is **off** updated value is 0. The default updated value is 1 [15].

"WavePad" by NCH software [16] was used to record the user audio sound command and F&D A520U 2.1computer multimedia speaker was used to play that recorded audio sound. A wave file with random frequencies was used as the added noise [17] and sound level meter TENMARS TM-102 was used to read measurements of all the sound levels.

III. Conclusion

The performance of the voice recognition system developed using Kinect V2 is evaluated in this study. In the first stage a stationary noise source was kept 3 m apart from the Kinect V2 sensor. The sound level measured at the point of Kinect V2 sensor is first set to 53.0 dB ± 0.5 dB. Then the accuracy of the voice command introduced at varied distance ranging from 1 m, 2 m, 3 m, and 4 m under the constant noise levels. According to results, the system recognizes the voice command with 95% accuracy when 4 m apart and increased the accuracy when distance to the sensor is decreased as illustrated in Figure 5. As the next step the performance of the system was evaluated with increasing levels of background noise which varied from 53 dB, 56 dB, to 58 dB. The accuracy of recognizing the voice command decreased with increasing background noise yet indicated highest accuracy at 4m distance when 53 dB background noise is present. In conclusion, the system can successfully detect the voice command even at noisy environments, thus can be used to introduce multiple controlling commands in the next stage.

In a study like this, it is possible to have many more future developments. As the next step we hope to train this system for each person, then user will only have to train his/her voice predefined times at the beginning. In that way we will be able to get a higher accuracy than the present accuracy of the system.

Authors wishes to acknowledge grants ASP/01/RE/SCI/2015/27 and ASP/01/RE/SCI/2016/25 supported for this project.

IV. References

[1] M. Hansson, E. Johansson, J. Lundberg and M. Otterberg, "Voice-operated Home Automation," Chalmers University of Technology, Bachelor of Science Thesis in Computer Science, June 2013.

[2] C. Reinisch ,"Wireless Communication in Home and Building Automation", Master thesis, Vienniauniversity of technology, Feb 2007.

[3] N. S. Liang, L.C. Fu and C. L. Wu, "An integrated, flexible, and Internet-based control architecture for home automation System in the Internet era," Proceedings ICRA, IEEE International, Conference on Robotics and Automation, vol. 2, pp. 1101 –1106, 2002.

[4] N. Sriskanthan and T. Karande, "Bluetooth Based Home Automation Systems," Journal of Microprocessors and Microsystems, vol. 26, pp. 281-289, 2002.

[5] K. Tan, T. Lee and C. Y. Soh, "Internet-Based Monitoring of Distributed Control Systems-An Undergraduate Experiment," IEEE Transactions on Education, vol. 45, No. 2, May 2002.

[6] N. Swamy, O. Kuljaca and F. Lewis, "Internet-Based Educational Control Systems Lab Using Net-meeting," IEEE Transaction on Education, vol. 45, No. 2, pp. 145-151, May 2002.

[7] P. Lin and H. Broberg, "HVAC Applications," IEEE Industry Applications Magazine," pp. 49-54, January 2002.

[8] K. G. H. A. Somasiri and W. K. I. L. Wanniarachchi, "Home Automation System, 6th Research Symposiu 2016," UvaWellassa University, ISBN 978-955-0481-09-5

[9] P. Gupta, "Human Voice Controlled Home Appliances," IJARCCE, vol. 4, Issue 5, pp 2319-5940 ISSN (Print), May 2015.

[10] S. Aravind, A. Mathew, S. Sugathan, R. Rajasre and S. Priyalakshmi, "Voice Controlled Smart Home", vol. 5, Issue 1,pp 272-276, January 2015.

[11] Microsoft. [2016, September 14]. Audio Stream [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131026.aspx

[12] Team of Principal Technologists. [2016, September 14]. Kinect SDK: Beamforming [Online]. Available: http://www.contentmaster.com/kinect/kinect-sdk-beamforming

[13] Microsoft. (2016, September 26). AboutSpeechRecognition Grammars (Microsoft.Speech) [Online]. Available: https://msdn.microsoft.com/en-us/library/hh378438(v=office.14).aspx

[14] Arduino (2016, September 04). Arduino Uno [Online]. Avalible: http://arduino.cc/en/Main/arduinoBoardUno

[15] Microsoft. (2016, September 26). SpeechRecognitionEngine.UpdateRecognizerSetting Method [Online]. Available: https://msdn.microsoft.com/en-us/library/system.speech.recognition.speechrecognitionengine.updaterecognizersetting(v=vs.110).aspx