# Home Automation by Speech Recognition

**Shilpa Srivastava, Ashutosh Kumar Singh, Sanjay Kumar Nayak**

***Abstract-- The speech control is now most important feature of a smart home. In this paper, we projected voice command module that is used to enable the user for a hands-free interaction between smart home and himself. We presented mainly three components that is required for a simple and an efficient control of smart home device(s). The wake-up-word parts allows the actual speech command processing. The voice recognition part maps the spoken voice command to text and then Voice Control Interface passes that text into an appropriate JSON format for the home automation. We evaluate every possibility of using a voice control module in the smart home by distinctly analyzing each and every component of module.***

***Keyword:Smart home, deep neural network, sensors, wake-up-word, speech recognition.***

## I. INTRODUCTION

The demography of the total population demonstrates a pattern that the older population worldwide is expanding quickly because of the expansion of the normal life expectancy of individuals. Thinking about and supporting this developing population is a worry for governments and countries around the world. Home automation is one of the real developing enterprises that can change the manner in which individuals live. A portion of these home automation frameworks focus on those looking for extravagance and complex home mechanization stages; others focus on those with exceptional needs like the old and the debilitated. The point of the announced Wireless Home Automation System (WHAS) is to furnish those with uncommon needs with a framework that can react to voice directions and control the on/off status of electrical gadgets, for example, lights, fans, TV and so on, in the home. The framework ought to be sensibly modest, simple to arrange, and simple to run. Figure 1: uControl Home Security, Monitoring and Automation (SMA). There have been a few business and research extends on brilliant homes and voice acknowledgment frameworks. Figure 1 demonstrates an incorporated stage for home security, observing and automation (SMA) from uControl . The framework is a 7-inch contact screen that can remotely be associated with security cautions and other home apparatuses. The home mechanization through this framework requires holding and communicating with a huge board which limitations the physical developments of the client. Another prominent economically accessible framework for home automation is from Home Automated Living (HAL). HAL programming taps the intensity of a current PC to control the home. It gives discourse order interface. A major bit of leeway of this framework is it can send directions everywhere throughout the house utilizing the current roadway of electrical wires inside the home's dividers. No new wires implies HAL is simple and cheap to introduce. Nonetheless, a large portion of these items sold in the market are intensely evaluated and regularly require huge home make over.


**Figure 1: "uControl Home Security, Monitoring and Automation (SMA)"**

The smart home represents an automated system in which various types of sensors and intelligent devices connect and work together to provide an efficient function to service an improved quality life for a comfortable house living. A smart home can be seen as the promising way of a supportive independent living providing an in home assistance that is especially important to the people having some disabilities [1].

The speech recognition system is considered to be the bridge for the betterment of the more natural human to computer interaction [2]. The recent advancement in the automated speech recognition (ASR) has made this technology one of essential features ina home automation(s) (HA) system [3][4]. Our mainobjective is to enable the voice/speech control ina home laboratory which can be then used to control huge number of devices like plugs, lights, dimmers, sensors, etc.

In the paper we present an architecture of a voice controlled module that has the role of translating a spoken command to a text and after that mapping them back to an appropriate actions in a HA system. The module constantly listens and then process sounds from environment to ensurea hands free interaction. To decreasetotalnumber of cases in which the action in a HA system is activated without actual person intent, we presenteda wake-up-word (WUW) detection device.

**Revised Manuscript Received on September 14, 2019.**

**Dr.Shilpa Srivastava**, Department of Information Technology, Noida Institute of Engineering and Technology, Noida, U.P. India.(Email: researchnietip@gmail.com)

**Ashutosh Kumar Singh**, Department of Electronics & Communication Engineering, Noida Institute of Engineering and Technology, Noida, U.P. India.(Email: researchnietip@gmail.com)

**Sanjay Kumar Nayak**, Department of Computer Science & Engineering, Noida Institute of Engineering and Technology, Noida, U.P. India.(Email: researchnietip@gmail.com)

*System Architecture*

In this part, we described an architecture of proposed voice controlled module. Proposed module consists of a Wake-Up-Word recognition module, a speech recognition engine and a voice command interface as shown in the Figure 2.

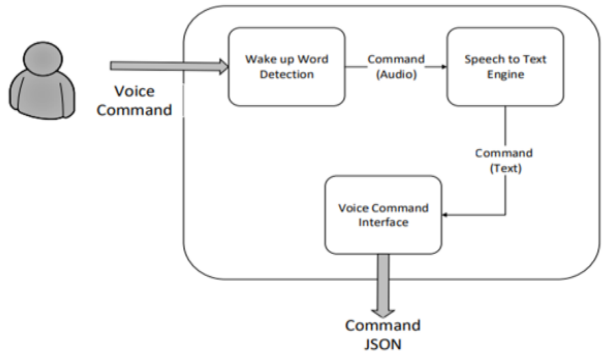


**Figure 2: A voice controlled module**

Firstly, there will be a brief introduction of deep neural network architecture of our WUW followed by description of each and every component.

*Deep Neural Network Architecture*

Due to the capabilityof the DNN to handle consecutive information nearlyevery end-to-end voice recognition system uses some neural network in its pipeline [5]–[8]. Inside RNN, an internal representation of the dynamic voice features is formed by feeding a low-level acoustic features into a hidden layer with a recurrent features from history. As there is an improvement of the conventional RNN that is a Bidirectional Recurrent Neural-Network(s) (BRNN) that can make the use of a future context. The data is then processed with both directions using two separate hidden layer by removing state neurons of the regular RNN in the part that is more responsible for anoptimistic time direction and part for a negative time direction that are then used for fed forwards to same output layer as shown in the Figure 3.

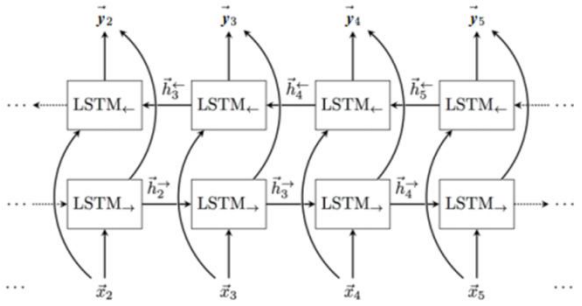


**Figure 3: Bidirectional long-short-term memory cell(s)**

Commonly, it is more difficult to train a RNN with the generally used activation functions like tanh, RELU, sigmoid, etc.becauseof exploding and the vanishing gradient [9]. One solution is to implanta memory structure for e.g. a long-short-term memory cells (LSTM) [10]. A structure of the LSTM memory cell is shown in the Figure 4.

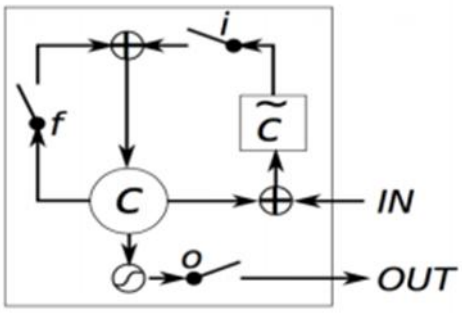


**Figure 4: A simple architecture of LSTM**

*Wake-up Word Detection*

Speech command module constantly receives and processes the sound. The main goal of the Wake-up-Word (WUW) detection module is to reduce additional number of case in which the ASR preforms a Speech-to-Text (SST) operations at same time for minimizing the unwanted triggers of the Voice Command Interface (VCI). Consequently, WUW module can be used as a constant identification method of a predefined wake-up-word whereasit rejects all the other words, noises and sounds.

## II. METHODOLOGY & RESULTS

The methodology of Voice Controlled module is prepared by analyzing each and every component distinctly. Before start of evaluation we asked 25 participants to read text from given paper that contains about 450 sentences of line in which 225 were automation commands. Every command line is then preceded by a WUW keyword “Alice”. Every recordings are then altered by a corrupting clean sounds using a simulator, collecting and then summingdifferent degrees of the noise and the reverberation from day to day life noisy recording so that the overall SNR lies between 1db and 40db with an average SNR of 14 db. The recorded data is about 6 hours long and this will be used as our test data set.

In this paper publically available recording(s) of the transcribed English voice are used as training sets for both the WUW and the ASR components. The training corpus consists of 900 hours of voice available as the part of a LibriSpeech dataset [11].

The experimental results for the WUW are given in regarding the false alarm per hour that is a false alarm frequency (FAF) that is defined as an incorrect detection of the keyword per hour. The false Alarm (FA) rate is then calculated by dividing number of the false positives by total number of the examples. The false alarm directs that the system is also activated even when user did not want to do so. The reported FAF of keyword spotter is 0.7. The evaluated results for the ASR component are given in the terms of a word error rate (WER). The reported average

WER of the LibriSpeeech corpus is about 21.5% and 32% when recorded test data set is used. Accuracy of the Command Interface lies most on the accuracy of the speech to text conversion, and then it can be achieved to 100% if command follows the grammar rules.

## III. CONCLUSION

In this paper, the presented architecture of the speech command module enables the voice control in the home lab. The main intent was to build the lightweight unit that can operate on the low power embedded devices such as Raspberry Pi. The presented WUW system enables user(s) to activate the VCM by using speech instead of the manual activation. The ASR was designed to allow an offline voice to text processing for matter of privacy. The Unstructured text from an ASR is then parsed by a Voice Command Interface and then converted into an appropriate JSON directive.

## REFERENCES

1. L. C. De Silva, C. Morikawa, and I. M. Petra, "State of the art of smart homes," Eng. Appl. Artif. Intell., 2012.
2. J. Picone, "Institute for Signal and Information Processing Fundamentals of Speech Recognition: A Short Course," Processing, 2008.
3. T. Giannakopoulos, N. A. Tatlas, T. Ganchev, and I. Potamitis, "A practical, real-time speech-driven home automation front-end," IEEE Trans. Consum. Electron., 2005.
4. I. V. McLoughlin and H. R. Sharifzadeh, "Speech recognition engine adaptions for smart home dialogues," in 2007 6th International Conference on Information, Communications and Signal Processing, ICICS, 2007.
5. A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2013.
6. A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks. 2012.
7. O. Vinyals, S. V. Ravuri, and D. Povey, "Revisiting recurrent neural networks for robust ASR," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2012.
8. J. Li, H. Zhang, X. Cai, and B. Xu, "Towards end-to-end speech recognition for Chinese Mandarin using long short-term memory recurrent neural networks," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015.
9. I. Sutskever, J. Martens, and G. E. Hinton, "Generating Text with Recurrent Neural Networks," Auk, 2011.
10. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., 1997.
11. V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2015.