VOL. 11, NO. 9, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A PROTOTYPE OF LOW COST IMPLEMENTATION OF AN INTELLIGENT HOME AUTOMATION SYSTEM

B. Bharathi, Simhadri Mahesh Kumar
Department of Computer Science and Engineering, Sathyabama University, Chennai, India
E-Mail: bharathivaradhu@gmail.com

ABSTRACT

Home Automation Systems have started gaining greater importance. The need of cognitive support for people with physical disabilities is the need of the day.The proposed idea of an Intelligent Home Automation System (VAHAS) is a simple way to control all electrical and electronic equipments at home using one's voice. All electrical equipments like fans, lights, television, air conditioners, etc are interfaced with the main system installed with the proposed software deployed on Linuz environment. The interfacing is brought about by RF enabled switches. The system is designed so that it responds to commands given by user. The system intelligently changes state of the equipments by activating and deactivting them without human intervention. This prototype when tested with a sample of 60 voice commands resulted in an average of 85% goodness. Extendibility is given utmost importance and customized functions can be added later.

Keywords: prototype, home automation system, cost effective, intelligent systems, voice activated systems.

INTRODUCTION

The wide applications of smart phones and other gadgets have increased the popularity of home automation. The voice is one of the most natural means to communicate and the intention is to bring that simplicity and elegance to the home by means of intelligent, voice based control system. Computers have evolved from Command Line Interfaces to Graphical User Interface and the next innovation in the user interface system has to be based on voice.

The commands are natural English commands and the focus is to understand those language commands. Natural Language Processing is a domain of computer science, soft computing, artificial intelligence and linguistics related with human, computer interactions. NLP is from the domain of human-computer interaction. Many challenges in NLP involve the understanding of natural languages, creating the ability of computers to decipher the meaning of inputs as human or natural language, and also involve generation of natural languages.

Statistical machine learning is the basic form modern Natural Language Processing algorithms. The paradigm of machine learning is different from that of most prior attempts at language processing. Prior implementations of language-processing tasks typically involved the direct hand coding of large sets of rules. Machine learning algorithms facilitate the continuous learning of rules from the real world examples. This is behaviour very specific to machine-learning in comparison to general learning algorithms. A corpus is a set of documents that have been hand-annotated with the correct values to be learned. [10]

Natural Language Processing applies many different classes of machine learning algorithms. These algorithms take a large set of "features" as the inputs. Some of the earliest-used algorithms, such as decision trees, produced systems of hard if-then rules similar to the systems of hand-written rules that were then common. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input feature.

Similar classes of systems diversely used in the market are analysed. The method of activating the automation system, the input signal used, (Table-1) the various language support provided are considered to indentify the requirements of the current system. [5]

A. Ubi

The Ubi plugs into a power outlet and connects to the world wide network through the local wireless internet connecting device. Integrated into the Ubi are two microphones, speakers, indicator lights and many sensors. Speech to text on the Ubi is a accomplished online and therefore the device requires internet for basic usage. Also the system is language dependent in the sense that it only supports English.

B. Voicepod

The Voicepod is a device which could be placed anywhere in your house preferably close to the person voicing the commands. It contains a high sensitivity microphone which can sense sound toupto 200 square feet. Color Light Emitting Diodes are present for visual feedback. The Voicepod supports English and Spanish but adding anymore locale specific languages is not a trivial task, and is thereby not a very scalable system. Again, this device requires a stable wireless network connection for its speech to text conversion capability.

ARPN Journal of Engineering and Applied Sciences

Table-1. A Comparison table of existing systems.

Factors	Estimated price	NLP input	Dynamic learning	Multiple language support	Security	Google search capability
VAHAS	$350	Yes	Yes	Yes	Yes	Yes
Ubi	$300	Yes	No	No	No	Yes
Voicepod	$650	No	No	Yes	Yes	Yes
Ivee Sleek	$230	Yes	No	Yes	No	No
mControl	$175	No	No	Yes	No	No
HAL2000	$665	No	No	No	No	Yes

C. Ivee sleek

Ivee Sleek is a voice-activated assistant that can answer questions, respond to commands, and connect to other smart home devices. Ivee is able to answer questions across 30 different categories like weather, time, etc.

D. mControl

mControl is a home automation device which could be used to automate every device, from lamps to thermostats to Heating, Ventilation and Air Conditioning systems. This device requires voice input in the form of predefined command sets and is not capable of natural language processing. It does not include the capability of executing searches online (Google) and also lacks scalability in the case of local language support, being restricted to English and Spanish.

E. HAL2000

HAL2000 is an operating system that enables you to control all the devices of your home, including lights, devices, home theatre, etc from anywhere in the world using your voice. It requires a fully fledged computer and does not process natural language. Further it is limited in its language support.

PROBLEM DEFINITION

Most current systems lack capabilities like multiple language support, usually being restricted to one or two. In a country like India where several different languages and their dialects are spoken in different parts, scalability is a must. Adding Indian language support in most of the devices mentioned above ranges from difficult to impossible.

In areas that lack deep penetration of internet bandwidth, a requirement of constant internet connectivity can also become a hurdle to users. All of the devices mentioned above process speech to text by means of web API calls, thereby being rendered unusable when they are offline.[3]

Most of the devices require the user to remember a set of predefined commands or instructions. This is because they do not support the capability of understanding natural language. The learning curve associated with using a device that is capable of processing natural language input is exponentially lower than of one that is not. This further increases the barrier that a user may face whilst trying to use one of these devices. [4]

This paper proposes a device called the Voice Activated Home Automation System (VAHAS) that aims to overcome the constraints and shortcomings of the existing devices that were surveyed. These deficiencies and the means to overcome them are as follows:

- The VAHAS is an offline system. The speech to text conversion process of the system does not require any active network or internet connection. This reinforces in the reduction of threats and penetration issues of insufficient bandwidth areas.
- Artificial intelligence with Natural language processing is the next human interface mechanism, and trumps voice command input. The device aims at ease of use than any of its predecessors by effectively leveraging the power of natural language.
- The concept of dynamic learning is presented in the device discussed in this paper, enabling the device to be personalised through use and making it capable of learning new tasks and intents.

SYSTEM DESIGN

The main device is designed around a Raspberry Pi running a voice recognition engine, a speech engine, natural language processing capable software and artificial intelligence driven algorithms to facilitate natural interaction with the system. The speech engine enables the system to speak back to the user, while the concept of positive reinforcement shall enable learning.

The switch is a microcontroller based, RF enabled board, designed to be modular and integration friendly with most commercial electrical switch designs. These not only allow control of power outlets via the main device, but also include capacitive touch capabilities, for secondary input in case of network failures. Other features include visual feedback and encryption for security.

The Raspberry Pi acts as the Server in this insistence, while the client is the Arduino board. This setup is replicated in the prototype as well. To connect the raspberry pi and the Arduino board, 2 Xbee series 2 modules with antennae are used and they establish a wireless connection. When the instruction is processed by the raspberry pi, it is sent to the client and it is executed. The instructions are performed by the Arduino using a programming language called Racket. [6]

The basic test for the system is done using the server-client mechanism. When a specific colour is specified by the user, the Raspberry Pi processes the request and sends the information through Zigbee protocol to the client. Now, on the client side the colour of the R.G.B. l.e.d turns into the colour specified and the server-client connection is successful. [7]

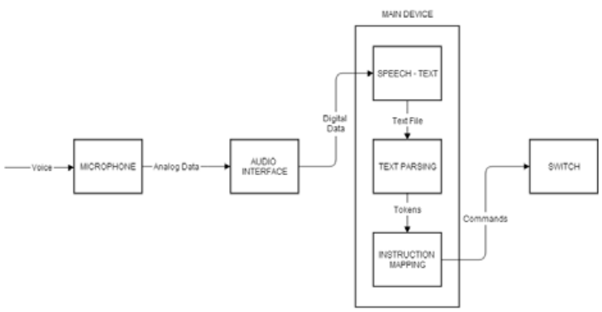

Figure-2. Interworking of algorithms.

A. Listener

The listener is the first stage of the entire process.It waits for the audio signal. Once it identifies the audio signal, it differentiates between voice commands and noise. This is the pocket-sphinx application that uses continuous listening with silence filtering to automatically segment a continuous stream of audio input into utterances that are then decoded. Each utterance is ended when a silence segment of at least 100 milliseconds is recognized. All of the audio received in this module is sent to the Noise Removal module of this system.

B. Noise removal

Speech to text conversion is the initial state of the entire process. Removing noise is an important part to differentiate noise from coherent words and given instructions. Pocketsphinx is completely integrated with the noise removal algorithm to increase the efficiency of the speech-text operation. Machinery sounds, noise from traffic or household are inserted during the recording or receiving of speech into the listener module is considered as unwanted noise added. Deduction of noise from speech patterns and syllables of the user providing the input as voice commands, makes the noise removal algorithm an important pre-processing step. The data is sent to the next module after the noise removal process. Hidden Markov Models module receives the noise removed audio data as input for the next step of processing.

C. Hidden Markov models

The next phase of the system is the speech to text conversion. In this section, the Hidden Markov Models (HMMs) are used extensively. This is because when an audio signal of a word is received by the Raspberry Pi, it is broken down into its individual syllables by this module, the Hidden Markov Models. The word is understood using HMMs. There usage is justified because of a simple and effective framework that can be used for modelling time-varying spectral vector sequences. The way a user converses and stresses on different syllables always have slight variance every time a word is uttered. The HMMs accounts these differences by taking into account the stress on each individual syllable. It then converts them into a specific word by referring to the words in the corpus. [9]

D. Viterbi algorithm

The dynamic programming algorithm used called the Viterbi algorithm is used to find the most likely sequence of hidden states called the Viterbi path. It results in a sequence of observed events in the context of Markov information sources and hidden Markov models.

E. Corpus

Corpus facilitates the processing of NLP in this system. A Corpus or text corpus is a large and structured set of text. The corpus is a collection of all vocabulary that the system needs to understand. If there is any input word mentioned by the user which is not present in the corpus, then the system does not respond [8].

In this application, whitelists are used to enable only a certain number of switches or clients to connect to the server (Raspberry Pi). Every new switch synced to the main device in a room is whitelisted for future switching on/off using the Pi. Client IDs can also be removed from the whitelist by instructing the Pi to delete a particular switch from the synced devices.

In this system we use Xbee shields to wirelessly connect the Raspberry Pi to the Arduino. Therefore every Xbee has a unique ID which is recorded in the Xbee shield connected to the Raspberry Pi. AXbee client which isn't whitelisted on the server cannot be connected to carry out the operations.

RESULTS AND DISCUSSIONS

The voice commands of male and female participants were considered for test runs. An average of 90% average success rate was noted. The test runs for male and female participants were conducted 30 times out of which 27 and 26 of them were declared successful. The end result averaged 90% and 86.67% respectively for male and female participants.

The speech to text based system is a phonetic based system and its accuracy is based on multiple factors. [2]

- The mismatch of the acoustic model. To verify this hypothesis a language model needs to be constructed from the test database text. Such language model will be very good and must give a high accuracy. If

accuracy is still low, more work needs to be done on the acoustic model. The acoustic model adaptation can be used to improve accuracy.
- The mismatch of the language model. Separate language models can be created for specific languages which need to be decoded.
- The mismatch in the dictionary and the pronunciation of the words. In that case work must be done on the phonetic dictionary.

The speech to text system has been optimized for application on a real time system. It achieves 90-95% accuracy post usage.

Figure-3. Word error rate.

Figure-3 shows that the word error rate of speech to text conversion module. The word error rate increases at a high rate when the number of words in the corpus increases. The corpus used in the proposed system contains only up to 150 words ensuring a manageable word error rate.

Trigger

The main device is continuously listening or waiting of user commands as inputs. This leads to a vulnerability of even general conversations among people can be considered as inputs and be processed by the device or system.[1] A simple initiation or trigger to identify the start of the command to the system ensures that the system process only specific commands given to the device on purpose. The system remains in a sleep mode at all other periods.

Removal of time delays

In the initial testing phase the Development Kit provided multiple time delays that lead to a response time of almost 6-7 seconds. A real time system cannot have an untenable response time. The following steps were taken to ensure reduction in time delay.

- There was a 2 second time delay between processing a word during the speech to text phase of the system. This delay was reduced to 200 milliseconds ensuring that the system doesn't require an excessive amount of time identifying the difference between speech and noise.
- Most of the code was generic for a standard 64-bit operating system. This code was completely modelled into 32-bit specific code which would perform at its best on the raspberry pi. This is responsible in a 70-80% reduction in the time delay.
- The GUI of the raspberry pi on its operating system Raspbian was completely done away with. For debugging purposes an SSH connection was established onto a laptop and a Command Line Interface was used. It is now referred to as a headless Unit.

Multiple language scalability

At present the language model for English is available and is being used. Pre-built language models are easily available for French, Chinese, German, Russian etc. Other language models can also be built for local Indian languages provided corpus for each of these languages are built with perfection.

Because of this advantage, this system could be used with any language provided a corpus is written for the language and the system is programmed to enable that particular language commands.

CONCLUSIONS

The paper presents the implementation of a low cost and effective way of home automation system. The implementation of natural language input into a voice activated home automation system has great demand for future market. Natural language makes it easier for users to use this system thus making this more efficient and user-friendly.

REFERENCES

[1] Acero A. 1993. Acoustical and Environmental Robustness in Automatic Speech Recognition, Boston, MA, Kluwer Academic Publishers. pp. 125-131.

[2] Allen J. 1995. Natural Language Understanding, 2nd edition, Menlo Park CA. The Benjamin/Cummings Publishing Company. pp. 237-242.

[3] Balentine B. and D. Morgan. 1999. How to Build a Speech Recognition Application. Enterprise Integration Group. pp. 325-337.

[4] Bernsen N., H. Dybkjar and L. Dybkjar. 1998. Designing Interactive Speech Systems, Springer. pp. 182-187.

[5] C. Kim. 2007. Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. IEEE Transactions on Audio, Speech and Language Recognition. 3(2): 4104-4104.

[6] G. Doblinger. 1995. Computationally Efficient Speech Enhancement by Spectral Minina Tracking.In:

ARPN Journal of Engineering and Applied Sciences

the Proceedings of Euro Speech Conference. 1(1): 21-29.

[7] Daniel Jurafsky and James Martin. 2009. Speech and Language Processing 2nd Edition, Pearson. pp. 179-192.

[8] George Luger. 2009. Artificial Intelligence 6th Edition. Addison Wesley. pp. 108-113.

[9] Mark Gales and Steve Young. 2008. The application of Hidden Markov Model in Speech Recognition, Foundation and Trends in signal processing. 1(3): 195-304.

[10] B.Bharathi. 2015. Location Based Services with Location Centric Profiles. Contemporary Engineering Sciences. 8(6): 221-227.