RESEARCH ARTICLE OPEN ACCESS

Gesture & Speech Based Appliance Control

Dr. Sayleegharge, Prof. Manisha Joshi, Bhaumikdoshi, Varun Sharma, Anikettamhankar
Department of Electronics & Telecommunications Vivekanand Education Society's Institute of Technology Collector Colony, Chembur (E), Mumbai-74.

Abstract
This document explores the use of speech & gestures to control home appliances. Aiming at the aging population of the world and relieving them from their dependencies. The two approaches used to sail through the target are the MFCC approach for speech processing and the Identification of Characteristic Point Algorithm for gesture recognition. A barrier preventing wide adoption is that this audience can find controlling assistive technology difficult, as they are less dexterous and computer literate. Our results hope to provide a more natural and intuitive interface to help bridge the gap between technology and elderly users.

I. INTRODUCTION

The demography of the world population shows a trend that the elderly population worldwide is increasing rapidly as a result of the increase of the average life expectancy of people. Caring for and supporting this growing population is a concern for governments and nations around the globe. Home automation is one of the major growing industries that can change the way people live. Some of these home automation systems target those seeking luxury and sophisticated home automation platforms; others target those with special needs like the elderly and the disabled.Aim is to build an embedded device which can single-handedly control all the household appliances. This device will be interfaced with a microcontroller which will eventually control all the electronic devices in the house. Hoping to provide those with special needs with a system that can respond to voice commands and control the on/off status of electrical devices, such as lamps, fans, television etc., in the home. The system should be reasonably cheap, easy to configure, and easy to run [1]

1. Speech Processing

Speech recognition applications are becoming more and more useful nowadays. Various interactive speech aware applications are available in the market. But they are usually meant for and executed on the traditional general-purpose computers. With growth in the needs for embedded computing and the demand for emerging embedded platforms, it is required that the speech recognition systems (SRS) are available on them too. PDAs and other handheld devices are becoming more and more powerful and affordable as well. It has become possible to run multimedia on these devices. Speech recognition systems emerge as efficient alternatives for such devices where typing becomes difficult attributed to their small screen limitations

2. Gesture Control

Gesture recognition is a topic in computer science and language technology with the goal for interpreting all the human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques

This paper starts with working of the speech and gesture system as explained in section II, followed by the main algorithms used in the system explained in section III. Going on the design i.e. how actually will it be implemented is given in section IV, section V concludes the paper, whereas section VI elaborates the future aspects of the project.

II. WORKING

1. Speaker identification

a) MFCC approach:

A block diagram of the structure of an MFCC processor is as shown in Fig 1. The speech input is typically recorded at a sampling rate above 10000 Hz. This sampling frequency was chosen to minimize the effects of *aliasing* in the analog-to-digital conversion. These sampled signals can capture all frequencies up to 5 kHz, which cover most energy of sounds that are generated by humans. The main purpose of the MFCC processor is to mimic the behavior of the human ears. In addition, rather than

the speech waveforms themselves, MFCC's are shown to be less susceptible to mentioned variations.

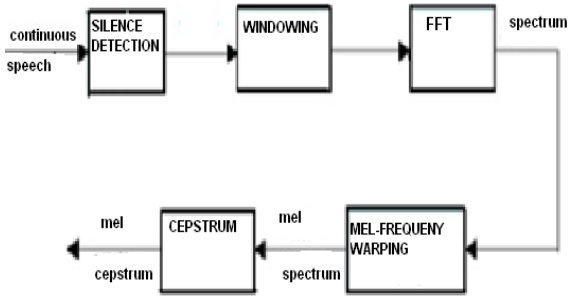

Fig. 1.MFCC Approach

Store the speech signal as a 10000 sample vector. It was observed that the actual uttered speech eliminating the static portions came up to about 2500 samples, so, by using a simple threshold technique the silence detection was carried out to extract the actual uttered speech .It is clear that the result was a voice based biometric system capable of recognizing isolated words. As experiments revealed almost all the isolated words were uttered within 2500 samples. But, when passed, this speech signal through a MFCC processor, it spilt this up in the time domain by using overlapping windows each with about 250 samples. Thus when converted into the frequency domain there were only 250 spectrum values under each window. This implied that converting it to the Mel scale would be redundant as the Mel scale is linear till 1000 Hz. So, elimination of the block which did the Mel warping.Direct use of overlapping triangular windows in the frequency domain.By obtaining the energy within each triangular window, followed by the DCT of their logarithms to achieve good compaction within a small number of coefficients as described by the MFCC approach. This algorithm however, has a drawback. As explained earlier the key to this approach is using the energies within each triangular window, however, this may not be the best approach as was discovered. It was seen from the experiments that because of the prominence given to energy, this approach failed to recognize the same word uttered with different energy. Also, as this takes the summation of the energy within each triangular window it would essentially give the same value of energy irrespective of whether the spectrum peaks at one particular frequency and falls to lower values around it or whether it has an equal spread within the window. It is an efficient method at lower energies.[2]

b) Training and testing:

The most common approaches to voice recognition can be divided into two classes: "template matching" and "feature analysis". Template matching is the simplest technique and has the highest accuracy when used properly, but it also suffers from the most limitations. As with any approach to voice recognition, the first step is for the user to speak a word or phrase into a microphone. The electrical signal from the microphone is digitized by an "analog-to-digital (A/D) converter", and is stored in memory. To determine the "meaning" of this voice input, the computer attempts to match the input with a digitized voice sample, or template that has a known meaning. This technique is a close analogy to the traditional command inputs from a keyboard. The program contains the input template, and attempts to match this template with the actual input using a simple conditional statement.Since each person's voice is different, the program cannot possibly contain a template for each potential user, so the program must first be "trained" with a new user's voice input before that user's voice can be recognized by the program. During a training session, the program displays a printed word or phrase, and the user speaks that word or phrase several times into a microphone. The program computes a statistical average of the multiple samples of the same word and stores the averaged sample as a template in a program data structure. With this approach to voice recognition, the program has a "vocabulary" that is limited to the words or phrases used in the training session, and its user base is also limited to those users who have trained the program. This type of system is known as "speaker dependent." It can have vocabularies on the order of a few hundred words and short phrases, and recognition accuracy can be about 98 percent.

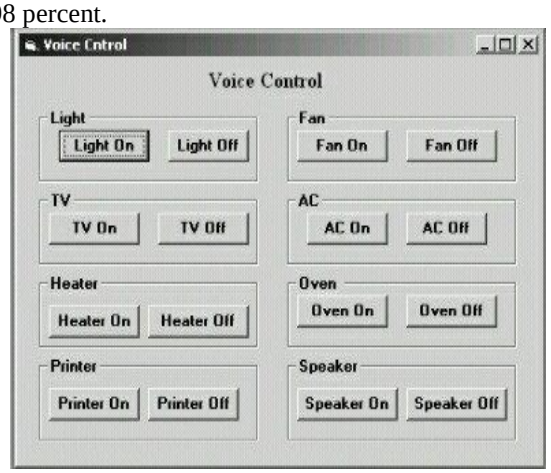

Fig. 2. GUI of voice control

The voice control window shown in fig.2 will provide the opportunity to control home appliances with the help of Voice command. The user needs only to utter the words with in the Command button to turn ON/OFF the specific device. Suppose, anybody wants to turn on the light, he needs only to utter the word "Light On".

2. Gesture control

By using a software that controls the cursor and does the clicks just by slight movements of thehand. The software provides real-time visual tracking and translates this into mouse controls. In addition the software is capable to interpret gestures as well as map these actions in keyboard events. All that is needed to be done is a training and testing exercise for the software. The software gets trained for certain gestures assign each gesture to a specific appliance. After which the web cam is turned after clicking on the gesture tab on the GUI that is built. The user gives some gesture input to the software which then will compare it with the previously taken samples. Accordingly the appliance will turn on or off[4] .

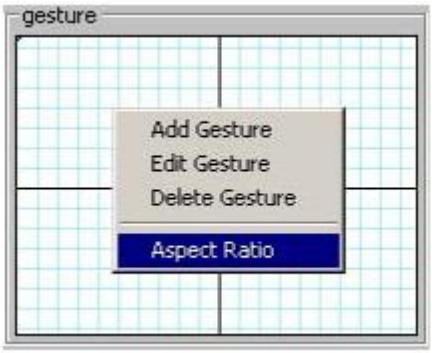

Fig. 3. Adding a gesture

In the gesture window you also have a popup menu as shown in Fig.2 from where you can either add a new gesture pattern or also set the aspect ratio for that windows as well.

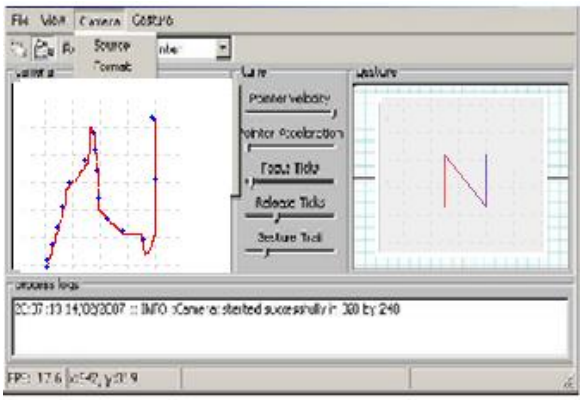

Fig. 4 .Training

After the gesture is done the trail of the gesture as shown in Fig.3 is used, by defining the length the trail left by the cursor in order to achieve a certain mark with a time frame this completes the training. In the testing phase then the trail is matched and then the related appliance turned ON.

III. ALGORITHMS

1. Speech algorithm

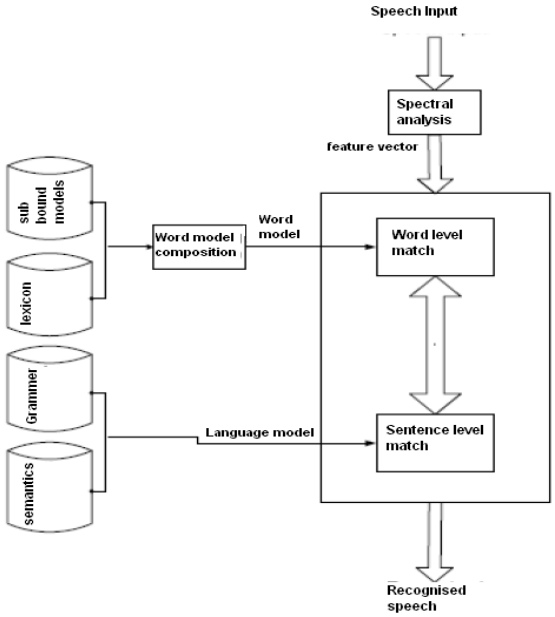

Fig.5.Flowchart of the speech recognition application

The application compares incoming speech with an obtainable predefined dictionary as shown in Fig.4. There are two main engines: Automatic Speech Recognition (ASR engine) and Text To Speech (TTS engine) .ASR implements the Fast Fourier Transform (FFT) to compute the spectrum of the speech input. Comparing the speech input with an existing database returns a string of the text being spoken. This string is represented by a control character that gets sent to the corresponding appliance's address. The designed graphical user interface (GUI) offers the user the choice of selecting the desired serial communication port as well as it provides a record of all the commands that have been recognised and executed. When designing the programme GUI, making it a user friendly application was a huge priority since the target clients need to avoid any possible complications in the system. Control characters corresponding to the recognised commands are then sent serially from the central controller module to the appliance control modules that are connected to the home appliances. [5]

2. Gesture algorithm

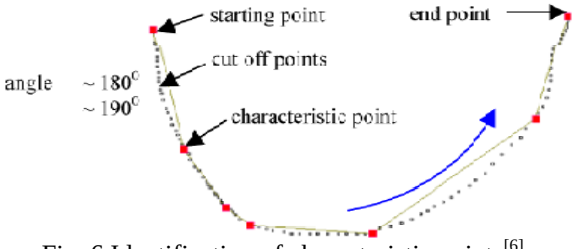

Fig. 6.Identification of characteristic points[6]

The characteristic points are looked upon that definethe significant changes in the direction of a line as shown in Fig.5, To do this, aheuristic algorithm of identification of the characteristic points is proposed. Itchecks whether the angle between the subsequent segments is close to 180°. If itis so, the segments are joined. Moreover, if some of the remaining points form atight group, all of them are removed except one.The significant change in the angle has beendefined as greater than 20°. The constraint of the minimal distance protectsagainst creation of spatial clusters of points. Aim is to keep only the pointsdefining the shape of a gesture.After removing all the points except the characteristic ones, the resultingnumber can be still different from the one requested by the classifier. In thesecond part of the algorithm two simple procedures are applied in order to: addsome points to the longest segments and remove some points having the shortestsegments joined respectively. This simple algorithm works remarkably well in thegiven domain

IV. DESIGN

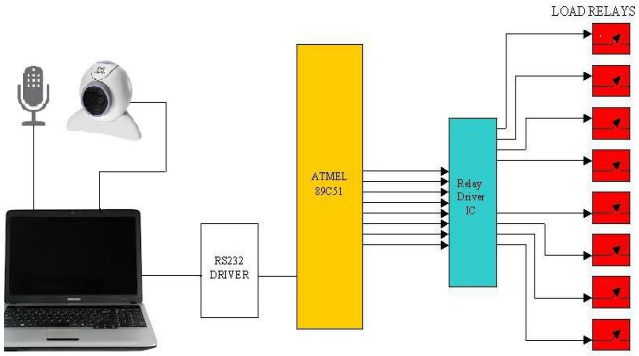

Fig. 7. Home appliances control through PC

This project is designed keeping a computer as the central controller. Input components will be a microphone & a webcam as the source for speech& gesture inputs respectively as shown in Fig.6.

The main components required in the hardware circuitry are:

- **ATMEL 89C51 microcontroller**:

AT89C51 is an 8-bit microcontroller and belongs to Atmel's 8051 family. ATMEL 89C51 has 4KB of Flash programmable and erasable read only memory (PEROM) and 128 bytes of RAM. It can be erased and program to a maximum of 1000 times.

In 40 pin AT89C51, there are four ports designated as P_1, P_2, P_3 and P_0. All these ports are 8-bit bi-directional ports, *i.e.*, they can be used as both input and output ports. Except P_0 which needs external pull-ups, rest of the ports have internal pull-ups.[7]

- **Relay Driver**:The ULN2003 is a monolithic IC consists of seven NPN Darlington transistor pairs with high voltage and current capability. It is commonly used for applications such as relay drivers, motor, display drivers and other high voltage current applications. It consists of common cathode clamp diodes for each NPN Darlington pair which makes this driver IC useful for switching inductive loads.[8]
- **RS232:** (Recommended Standard-232) A TIA/EIA standard for serial transmission between computers and peripheral devices (modem, mouse, etc.). Using a 25-pin DB-25 or 9-pin DB-9 connector, its normal cable limitation of 50 feet can be extended to several hundred feet with high-quality cable. RS-232 defines the purpose and signal timing for each of the 25 lines; however, many applications use less than a dozen.[9]
- **Load Relays:** these are connected to the appliances which are to be controlled.

TxD pin of serial port connects to RxD pin of controller via MAX232. And similarly, RxD pin of serial port connects to the TxD pin of controller through MAX232.MAX232 has two sets of line drivers for transferring and receiving data

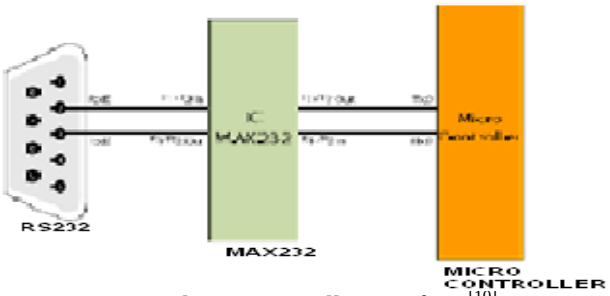

Fig. 8. RS232 and microcontroller interface.[10]

The line drivers used for transmission are called T1 and T2, whereas the line drivers for receiver are designated as R1 and R2. The connection of MAX232 with computer and the controller is shown in Fig.7.

The controller is then connected to a relay driver in this case ULN2003, which consists of seven NPN Darlington connected transistors in these arrays are well suited for driving lamps, relays, or printer hammers in a variety of industrial and consumer applications. Their high breakdown voltage and internal suppression diodes insure freedom from problems associated with inductive loads. Peak inrush currents to 500 mA permit them to drive incandescent lamps. The ULx2003A with a 2.7 kΩ series input resistor is well suited for systems' utilizing a 5.0 V TTL or CMOS Logic.This driver is then connected to the load relays which interfaced with the appliances.

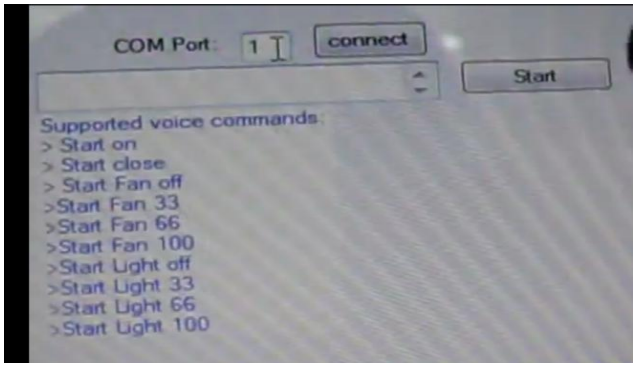

Fig. 9.Screenshot of the Application GUI.

Fig.8. shows the GUI of the speech interface .As we can see a varied number of home appliances are listedabove can be turned ON and OFF corresponding to the command.

V. CONCLUSION

Our aim is to build an embedded device which can single-handedly control all the household appliances. This device will be interfaced with a microcontroller which will eventually control all the electronic devices in the house. Such devices can be used by the elderly or disabled to operate the electronic appliances in the house efficiently. We aim to design the device with a natural and user-friendly interface. Major influencers as to how well an elderly/disabled person would be able to use and understand this technology are considered.

VI. FUTURE WORK

More features that can be added into this project & this future work may entail such features:

- Adding confirmation commands to the voice recognition system.
- Integrating variable control functions to improve the system versatility such as providing control commands other than ON/OFF commands. For example “Increase Temperature”, “Dim Lights” etc.
- Integration of GSM or mobile server to operate from a distance.
- Introduction of biometric systems in order to isolate each user & thereby provide more layers of security.

These features will make our prototype more efficient & secure. The design of the device will have a more user-friendly & natural interface. So, it’ll be more useful for the elderly & the disabled.

.

REFERENCES

[1] Y Bala Krishna et al, Int. J. Computer Technology & Applications, Vol 3 (1), 163-168
[2] Speaker Recognition Using MATLAB Available:http://www.scribd.com/doc/59159000/29/Matlab-Code-for-MFCC-approach
[3] Jim Baumann, “Voice Recognition” Available:http://www.hitl.washington.edu/scivw/EVE/I.D.2.d.VoiceRecognition.html
[4] Larryo.org/work/information/umouse
[5] Boukreev, K.: Mouse Gestures Recognition. CodeGuru, www.codeguru.com (December 2001).
[6] S.Nagendram ,Int. J. Computer Technology & Applications, Vol 3 (1), 163-168
[7] http://www.atmel.in/Images/doc0265.pdf
[8] http://www.electrosome.com/uln2003-high-voltage-current-driver/
[9] www.wikipedia.org
[10] http://www.engineersgarage.com/microcontroller/8051projects/interface-serialport-RS232-AT89C51-circuit