ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (383-391), Month: January - March 2015, Available at: www.researchpublish.com

Handling PC through Speech Recognition and Air Gesture

[1]Sameer Shewalkar, [2]Shoaib Ansari, [3]Masuma Mujawar, [4]Prof.Patil S.S

Abstract: **Speech recognition technology is one from the fast growing engineering technologies. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input. This project is designed and developed keeping that factor into mind, and a little effort is made to achieve this aim. Our project is capable to recognize the speech and convert the input audio into text; it also enables a user to perform operations such as "save, open, exit" a file by providing voice input. It also helps the user to open different system software such as opening Ms-paint, notepad and calculator. At the initial level effort is made to provide help for basic operations as discussed above, but the software can further be updated and enhanced in order to cover more operations.**

- **Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively.**
- **The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input and air gestures using the android phones.**

Keywords: **Speech Recognition, Air gesture.**

1. LITERATURE REVIEW

A) Live scribe smart pen is a pen like device that is used to track person's writing. This device requires a special finely dotted paper to monitor the movement of the pen. This pen can recognize alphabets and numbers, and it can download to PC. However, the dotted paper may not be always accessible, making quick note-taking difficult. Tablets PC's also suffer from this problem of quick accessibility. A series of application for the NOKIA N95 use the built in accelerometer.

B) The NiiMe project transformed the N95 phone into a Bluetooth PC mouse. The **PyAccele REMOTO-R** project developed a remote control for the Linux media player M player. By tilting the phone, the play, stop, volume, stop, fast-forward and rewind functions of the player are controlled, e.g. to guide a ball through a maze. Being able to write in air we believe is a more challenging problem than the ones in existing system.

C) The Logitech Air Mouse targets people who use computers as multimedia devices. The Air Mouse in air similar to a remote control. The accelerometer in this device and gyroscopes allow accurate linear .The presence of accelerometer and gyroscopes together allow accurate linear and rotational motion of the pointer on the screen. Unlike theAir Mouse, PhonePoint Pens are not equipped with gyroscopes and relies on just on accelerometers. Furthermore, the proposed phone-based pen does not have a screen on which one may see the pen movement in real time. The absence of gyroscopes and visual cues makes positioning of the pen a difficult problem.

D) A popular device capable of tracking hand movement is the **Wii remote (or "Wiimote")** used by **Nintendo Wii console .**The Wiimote uses a 3-axes accelerometer to infer forward and backward rapid movements. In addition, optical sensors aid in positioning, accurate pointing, and the rotation of the device relative to the ground. The optical sensor is

embedded on the Wiimote and relies on a fixed reference (and a sensor bar) centred on top of the gameplay screen. The "Wiimote" can be augmented with the "Wii Motion Plus", a pluggable device containing an integrated gyroscope. Using this feature rotational motion is captured. These three sensors – the accelerometer, the gyroscope, and the optical sensor- can reproduce motions similar to real arm-motion. The windows phone consists of only a (low-cost) accelerometer, and limited processing capabilities, in comparison to the Wii. Developing the pen on this platform entails a variety of new challenges. While noting down different stuff that you need to remember is not a hard thing to do at most times, there are certain situations when you just don't have the means or possibility to write something down, like those moments when you have your hands full of luggage and something pops up in your mind. Our proposed work allows cell phone users to write short notes in the air and send them automatically to server.

E) Sandip Agrawal and group from Duke University published a paper in June 2009 on application of mobile computing. In this paper they developed new software called Phone point pen for mobile device which worked as follows. The main work was to create a Phone Point Pen on the mobile phone device that could convert writing in the air to digital text and images **The Phone Point Pen** application first captures the user's hand movements, whether drawing a picture, jotting down numbers or writing a sentence. This software is able to translate the character written in English into a message that can be e-mailed and accessed either from a computer or mobile phone. In the research Pictures are saved as .jpg attachments that can be sent along with e-mails. As per literature published by Sandip Agrwal and group there existing system worked only to translate characters in English. Our proposed work is to recognize the characters in Marathi.

F) Air-gestures with 3D accelerometers. Gesture recognition has been widely studied through accelerometers, gyroscopes, vision based techniques, etc. Works that are closest to P3 include a sensor mote-based 4-character recognizer, a numeric digit recognizer with customized hardware and [15] uWave, a mobile phone based single-gesture recognizer. The first two works employ a highly capable accelerometer (around 100 samples/s). They use Principle Component Analysis (PCA), Hidden Markov Models (HMMs), and Dynamic Time Warping (DTW) algorithms, to achieve accuracies of 90 to 94%. However, the accuracy falls to 80% when the accelerometer is sampled at 40 samples/s. More importantly, the proposed systems are only able to write few numeric digits that do not require the user to reposition the pen within the same character. Geometric figures are also not viable because gestures are identified through pattern matching, and hence, the system does not compute the actual displacement and direction of motion. Phone Point Pen, on the other hand, tracks the user's hand movement, and develops methods for pen-repositioning, character transition, stroke-grammar, rotation avoidance, and character disambiguation.

G) *uWave* is a mature work that allows a user to gesture with mobile phones, enabling simple operations like gesture-based user authentication, opening/closing applications, etc. The authors attain an impressive 99% accuracy with 8 gestures and negligible training. While this is valuable for a number of interfacing applications, we emphasize that character recognition entails an additional set of problems. Specifically, gestures are significantly tolerant to error; as long as the errors repeat across all gestures, the gesture can be identified. In contrast, the Phone Point Pen requires a different approach to continuously track a more complicated motion of the hand.

2. SOFTWARE REQUIREMENTS SPECIFICATION (SRS)

2.1) INTRODUCTION:

The concept of speech recognition started somewhere in 1940s, practically the first speech recognition program was appeared in 1952 at the bell labs, that was about recognition of a digit in a noise free environment. 1940s and 1950s consider as the foundational period of the speech recognition technology, in this period work was done on the foundational paradigms of the speech recognition that is automation and information theoretic models. In the 1960's we were able to recognize small vocabularies (order of 10-100 words) of isolated words, based on simple acoustic-phonetic properties of speech sounds.

The key technologies that were developed during this decade were, filter banks and time normalization methods. In 1970s the medium vocabularies (order of 100-1000 words) using simple template-based, pattern recognition methods were recognized. In 1980s large vocabularies (1000-unlimited) were used and speech recognition problems based on statistical, with a large range of networks for handling language structures were addressed. The key invention of this era were

hidden markov model (HMM) and the stochastic language model, which together enabled powerful new methods for handling continuous speech recognition problem efficiently and with high performance.

2.1.1) Project scope:

This project has the speech recognizing and speech synthesizing capabilities though it is not a complete replacement of what we call a NOTEPAD but still a good text editor to be used through voice. This software also can open windows based software's such as Notepad, Ms-paint and more. This is also possible to control media, cursor and much more other operations by air gesture.

2.1.2) Users and Characteristics

The user of this system will be any user who as Android phone with much good handling capability to it and also able to operate computer system.

2.1.3) Operating Environment

Mobile application software will support Android phones with at least 2.1 versions.

Computer contains Windows xp, vista, and 07 operating system.

2.1.4) Design and Implementation Constraints

At least1 Android mobile device and one computer system needed to to show working model. Also system should have sufficient microphone quality and sensing capability to get process properly.

2.1.5) User Documentation

Training tool and User manual will be given as part of release.

2.1.6) Assumptions and Dependencies

The project will support only English language to get the appropriate command from the user. Mobile microphone quality will be good enough to access the user's voice command.

2.2) SYSTEM FEATURES:

External Interface Requirements:

2.2.1) User Interfaces:

1. Configuration page
2. Camera page

2.2.2) Hardware Interfaces:

Mobile application will get installed on mobile devices. These mobile devices should have Camera.Configued Computer.

2.2.3) Software Interfaces:

Minimum 256 MB RAM memories, preferably 512MB RAM

• 50MB free hard disk space.

• Good quality microphone for audio input.

Software Requirements:

• Java Software development kit 6.

• Web browser

• Audio drivers installed

For Mobiles:-

• Minimum 32MB RAM memory, preferably 64MB RAM

• 50MB free storage space.

• WAP enabled browser.

2.2.4) Communications Interfaces:

The interfacing consists of two parts:

• The interfacing should be correctly performed between the speech engine and the application for performing voice recognition.

• The deployment of the application on the mobiles involves interfacing between the mobile device and the application. The URLs opened will only be using the HTTP protocol.

Functional Requirements:

Table 2.1: Functional Requirement

Test	Input signal	Results	Discussions
1.Writing Text	(I) Keyboard	Printed output of the input signal on screen.	We Provided input through keyboard and by pressing each printable key and observed that it worked fine as expected.
	(II)Voice	Displayed the output of the input signal on screen.	While providing voice input to the software it recognized the spoken words in few attempts, this is due to noisy environment, variation in the voice and multiple user factor.
2. Running notepad commands (open, save, clear)	(I) Mouse/Keyboard	Properly functioning of open, save and clear to file were observed.	we properly run the notepad commands through mouse input and they worked properly fine.
	(II) Voice	Properly functioning of open, save and clear to file were observed.	we provide voice input to run the notepad commands and they work fine according to the expectations.
3. Running System commands (Calculator, Ms-Paint, Notepad)	(I)Voice	System soft wares were opened and Commands worked finely.	By providing voice input to the software for running system commands it worked fine and result in expectations, without repeating the commands twice or thrice.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (383-391), Month: January - March 2015, Available at: www.researchpublish.com

2.3) Other Non-functional Requirements:

3.3.1) Performance Requirement:

The application should perform accurately in a quiet environment. The performance should not deteriorate drastically even in the presence of some noise. The response time for recognizing a word should not be more than 2 seconds. The probability of misrecognition should be not more than 30% - 40%.

2.3.2) Safety Requirements:

NA

2.3.3) Software Quality Attributes:

The application should be platform independent.

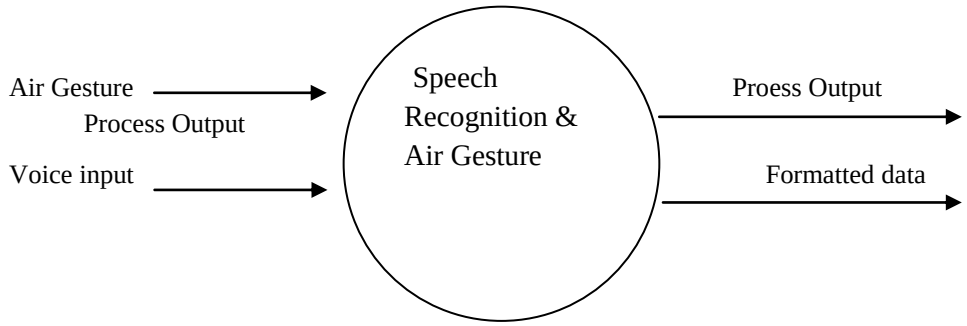
- The product will be readily available from any computer with access to the Internet.
- The product will be built using a modular scheme. Future modifications and additions should then be easy to implement, since only changes to certain modules will be necessary

2.4) Constraints

The speech engine to be used should be preferably an open source engine and should adequately satisfy the memory constraints of the device used. Owing to its platform independence Java should be used as the programming language.

2.5) ANALYSIS MODEL:

DFD:



2.5.1 –Level 0

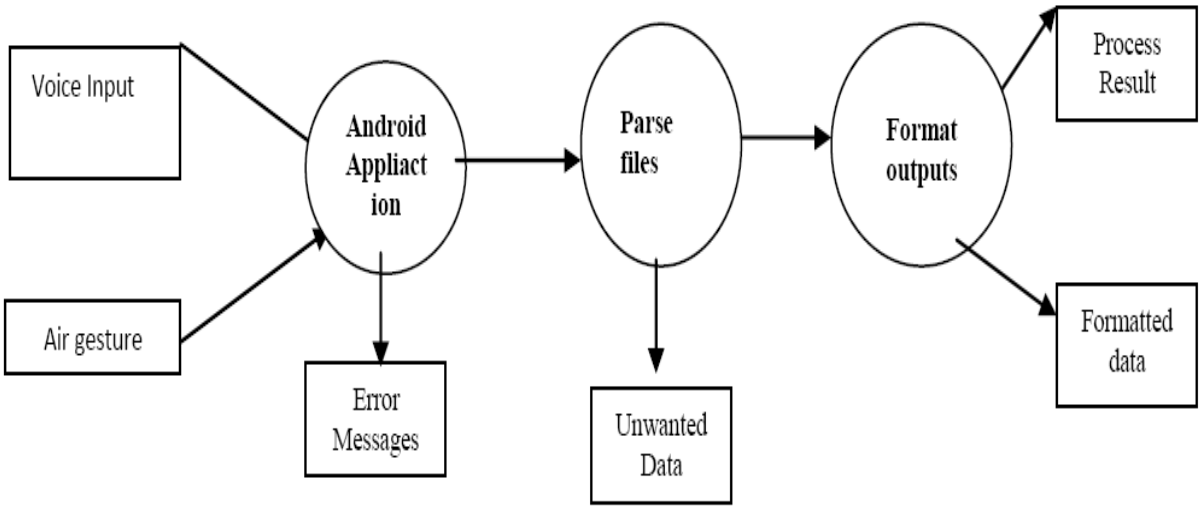
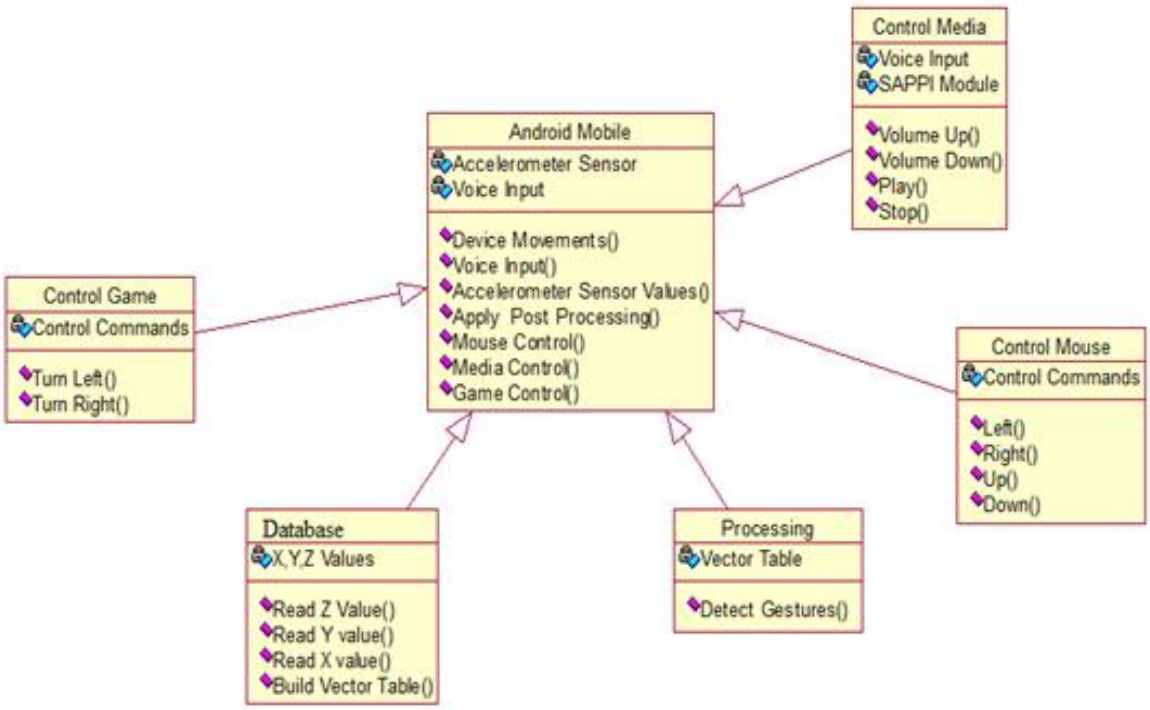


Figure 2.5.2 –Level 1

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (383-391), Month: January - March 2015, Available at: www.researchpublish.com



2.5.3: Class Diagram

The class diagram is the main building block of object oriented modelling. It is used both for general conceptual modelling of the systematics of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be used for data modelingThe classes in a class diagram represent both the main objects, interactions in the application and the classes to be programmed.

In the diagram, classes are represented with boxes which contain three parts:

- The top part contains the name of the class. It is printed in Bold, centered and the first letter capitalized.
- The middle part contains the attributes of the class. They are left aligned and the first letter is lower case.
- The bottom part gives the methods or operations the class can take or undertake. They are also left aligned and the first letter is lower case.

3. SYSTEM DESIGN

3.1) SYSTEM OVERVIEW & ARCHITECTURE:

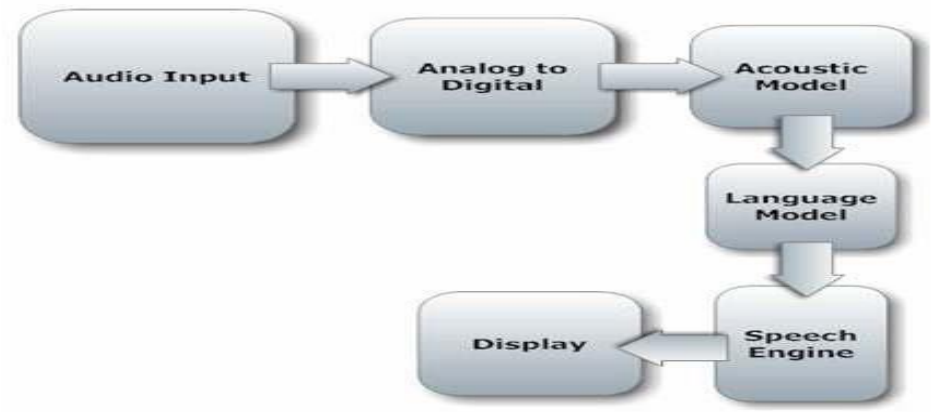


Figure 3.1: System Architecture

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (383-391), Month: January - March 2015, Available at: www.researchpublish.com

COMPONENTS OF SPEECH RECOGNITION SYSTEM:

1. Voice Input:

With the help of microphone audio is input to the system, the pc sound card produces the equivalent digital representation of received audio.

2. Digitization:

The process of converting the analog signal into a digital form is known as digitization , it involves the both sampling and quantization processes. Sampling is converting a continuous signal into discrete signal, while the process of approximating a continuous range of values is known as quantization.

3. Acoustic Model:

An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech. The software acoustic model breaks the words into the phonemes.

4. Language Model:

Language modeling is used in many natural language processing applications such as speech recognition tries to capture the properties of a language and to predict the next word in the speech sequence. The software language model compares the phonemes to words in its built in dictionary.

5. Speech Engine:

The job of speech recognition engine is to convert the input audio into text to accomplish this it uses all sorts of data, software algorithms and statistics. Its first operation is digitization as discussed earlier, that is to convert it into a suitable format for further processing. Once audio signal is in proper format it then searches the best match for it. It does this by considering the words it knows, once the signal is recognized it returns its corresponding text string.

3.2) AIR GESTURE:

Gestures are a form of nonverbal communication in which visible bodily actions are used to communicate important messages. Gestures include movement of the hands, face, or other part of the body. In the recent years sensing the hand gestures has become one of the applications of Mobile computing. Mobile Computing is a technology that allows transmission of data, via a computer, without having to be connected to a fixed physical link.

USE OF AIR GESTURE:

1. For Media Control
2. File Operation
3. Cursor Control

Figure 3.2: Air Gesture

Gestures are a form of nonverbal communication in which visible bodily actions are used to communicate important messages. Gestures include movement of the hands, face, or other part of the body. In the recent years sensing the hand gestures has become one of the applications of Mobile computing. Mobile Computing is a technology that allows transmission of data, via a computer, without having to be connected to a fixed physical link.

4. TECHNICAL SPECIFICATIONS

4.1) ADVANTAGES OF SYSTEM:

1. Able to write the text through both keyboard and voice input.
2. Voice recognition of different notepad commands such as open save and clear.
3. Open different windows soft wares, based on voice input.
4. Requires less consumption of time in writing text.
5. Provide significant help for the people with disabilities.
6. Lower operational costs.

4.2) DISADVANTAGES:

- Less accuracy
- Cost Effective
- Difficulty in noisy environment

4.3) HARDWARE REQUIREMENTS:

- 64 MB of RAM
- Microphone
- Sound card

Pentium 200 MHz processor

Best requirements:

- 1.6 GHz Processor
- 128 MB or more of RAM
- Sound cards with very clear signals
- High quality microphones

4.4) SOFTWARE REQUIREMENTS:

- Windows 07,xp OS
- Android version Jelly Bean, Kit-Kat

4.5) APPLICATIONS:

1) From medical perspective

People with disabilities can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands, in such cases speech recognition programs are much beneficial and they can use for operating computers. Speech recognition is used in deaf telephony, such as voicemail to text.

2) From military perspective

Speech recognition programs are important from military perspective; in Air Force speech recognition has definite potential for reducing pilot workload. Beside the Air force such Programs can also be trained to be used in helicopters,9 battle management and other applications.

3) From educational perspective

Individuals with learning disabilities who have problems with thought-to-paper communication (essentially they think of an idea but it is processed incorrectly causing it to end up differently on paper) can benefit from the software.

REFERENCES

[1] IEEE Paper on "Surround Sense Hand Gestures for Mobile Devices" by Prof. Mrs. S. Golait

[2] "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition". D. Jurafsky, J. Martin. 2000. ISBN: 0130950696

[3] Stephen Cook ""Speech Recognition HOWTO" April 19, 2002

[4] B.H. Juang & Lawrence R. Rabiner, "Automatic Speech Recognition – A Brief History of the Technology Development" 10/08/2004

[5] Ksenia Shalonova, "Automatic Speech Recognition" 07 DEC 2007 Source:http://www.cs. bris.ac.uk/ Teaching/ Resources/COMS12303/lectures/Ksenia_Shalonova- Speech_Recognition.pdf

[6] John Kirriemuir "Speech recognition technologies" March 30th2003.

APPENDIX – A

Acronyms:

- HMM :- HIDDEN MARKOV MODEL
- UML :- UNIFORM MODELING LANGUAGE