Design of Multi Lingual, Voice Signal Frequency Based Robotic Hand Control System

[1]Kartik Sharma, [2]Gianetan Singh Sekhon

[1]Student, [2]Asst Professor & In-Charge,Computer Engineering Section,
Yadavindra College of Engineering , Punjabi University Guru Kashi
Campus,Talwandi Sabo
[1]kartikavasthi @gmail.com, [2]gianetan @gmail.com

Abstract: *This Paper Presents the Design of Robotic Hand Control System which is multilingual and voice signal frequency based. MATLAB®'s Interactive programming Environment makes it an ideal tool for various specialized applications such as speech processing and signal processing. We have used a simple technique which matches the voice templates based upon their frequency values. The system makes use of the MATLAB®'s inbuilt functions for searching the database for the recorded voice templates and the robotic hand responds corresponding to the successful match.*

Index Terms: *Automatic Speech Recognition (ASR); Matlab®; Robotic Hand Control System*

1. Introduction

The theme of Social interaction and intelligence is important and interesting to an Artificial Intelligence and Robotics community. It is one of the challenging areas in Human-Robot Interaction (HRI). Speech recognition

technology is a great aid to admit the challenge and it is a prominent technology for Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) for the future. Humans are used to interact with Natural Language (NL) in the social context. This idea leads Roboticist to make NL interface through Speech for the HRI. Natural Language (NL) interface is now starting to appear in standard software application. It helps novices to easily interact with the standard software in HCI field. Its also encourage Roboticist to use Speech Recognition (SR) technology for the HRI [5]

Automatic Speaker recognition is the process of automatically recognizing who is speaking based on unique characteristics contained in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, database access services, information services, voice mail, and security control for confidential information areas, and remote access to computers [2]. Development of speaker identification systems began as early as the 1960s with exploration into voiceprint analysis, where characteristics of an individual's voice were thought to be able to characterize the uniqueness of an individual much like a fingerprint. The early systems had many flaws and research ensued to derive a more reliable method of predicting the correlation between two sets of speech utterances. Speaker identification research continues today under the realm of the field of digital signal processing where many advances have taken place in recent years [1, 3]

In this paper a basic simple multi lingual voice signal frequency recognition technique has been used to search the Database using the Matlab® and choose the most suitable match based on the pre defined and stored voice templates in database. This is a multilingual (Hindi & Punjabi) project so the commands given in Punjabi or in Hindi are matched and the Robotic Hand will move as commanded by the user.

2. System Design and Multi Lingual Approach:-

The main objective of this paper is to design and implement multilingul speech recognition system using Matlab®'s DSP toolbox, which is capable

of recognizing and responding to key words inputed in multi languages.This Multi Lingual voice signal frequency recognizer would be applicable on many Speech recognition devices [1]. In this research, basically we use Matlab® for matching the Frequency of spoken words in any language wheather in Hindi or Punjabi.

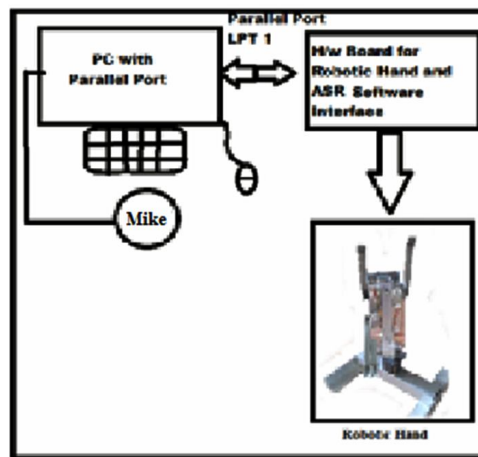*2.1)*   *Design of the System Model*

Figure1 System Model

The matching of the spoken words with the stored template/word is based upon their respective frequencies,if the frequency of the current voice is same as that of pre stored voice template.

In the figure 1 above the mike is attached to the PC, all the voice templates are recorded with the help of mike only, the Robotic hand is attached to the PC's Parallel Port (LPT1) through the Hardware circuit.

3. Methodology

There are two Components of our work

(1) System Software Development and

(2) Hardware Setup.

3.1 Development of System Software

During the software development phase of this project a database for

the voice samples or "voice fingerprints" was created in order to match these against the voice templates which are spoken by the user. A voice fingerprint represents the most basic, yet unique, features of a particular speakers' voice. In the first phase of our development we store the voice template of the hindi and punjabi language in the database. The next step is to define time frame for recording multiligual command words having duration (t) =40 seconds, frequency fs=8000 HZ, next is to record voice sample in hindi or Punjabi using "wavrecord" command.

Sr. No	Time Frame for Voice Sample			Functions Defined For Parallel Port(LPT1)
	Command in Hindi & Punjabi	English Language equivalent	Time Taken in Milliseconds	
1.	हाथ खू ले	Open hand	2000 To 7600	Matched=1
2.	अपने हाथ बंद करो	Close your hand	9600 To 12300	Matched=2
3.	अब आप कलाई दाहिने घुमाएँ	Now rotate your hand on the right hand side.	15800 To 19000	Matched=3
4.	ਹੁਣ ਪੰਜਾਬੀ ਵਿੱਚ ਸੱਮਝ ਦੇ ਨਾਲ ਆਪਣੇ ਹੱਥ ਆਪਣੇ ਖੱਬੇ ਪਾਸੇ ਹੱਥ ਦੇ ਵੱਲ ਘੁਮਾਉ	Now with the understanding in Punjabi please turn your hand tom your left hand side.	21400 To 35600	Matched=4
5.	ਅਤੇ ਹੁਣ ਜਿਆਦਾ ਸੱਮਝ ਦੇ ਨਾਲ ਸੱਜੇ ਹੱਥ ਦੇ ਵੱਲ ਆਪਣਾ ਹੱਥ ਘੁਮਾਉ	Now with More understanding rotate your hand to Right hand side.	21500 T0 25400	Matched=5
6.	ਕੀ ਤੁਹਾਡੇ ਹਾਰਡਵੇਯਰ ਅਤੇ ਸਾਫਟਵੇਯਰ ਠੀਕ ਚੰਗ ਨਾਲ ਕੱਮ ਕਰ ਰਿਹਾ ਹੈ ?	Is your hardware and software is working correctly?	22500 To 27500	Matched=6

Table 1 For Time Frame for Voice Samples

Table 1 Shows as the first command "हाथ खुले" Recorded in 2.0

milliSecond to 7.6 Milliseconds, if the Command Matched with the Input Command than the Function Matched=1; is executed and the correspondent work by the Robotic Hand is done.

3.3.1) First phase of development: *Recording of the First*

Hindi Keyword Command ("हाथ खुले"*)*

As shown in table 1 record these key words in database as .wav file having different time frames for each key word using these commands.

1) ***hindi1=wavrecord(t,fs);***

Here this command is used to record command word (hindi1) with parameters:

- Time frame "t=40000
- Frequency "fs=8000"

2) ***[hindi1_x hindi1_y]=find(temp1.hindi1 >0 .6);***

Here this command is used to take voice sample frequency amplitude above the value 0.6 and discard those voice sample frequencies which are less.

3) ***diff_hindi1 = max(hindi1_x) – min(hindi1_x);***

Here this command finds the difference between maximum and minimum value of speech sample   which is further used when we matched the Command with the Current voice sample

By the Given Previous method all the remaining keywords (hindi2, hindi3, punjabi1, punjabi2, punjabi3, start, stop) are recorded and then saved at a particular location of Computer Memory

4) ***save ('E:\Multilingual.mat',hindi1','hindi2,'hindi3' 'punjbai1' 'punjabi2' 'punjbai3' 'start' 'stop');***

Here this save Command is used to save the Given Spoken keywords in a articular memory location of Computer

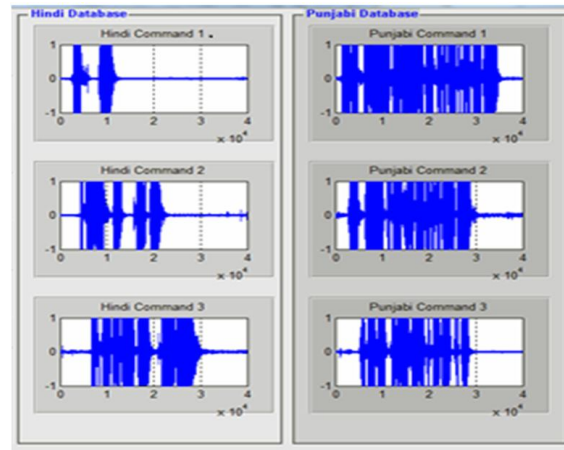This subplots the graph between time and amplitude axis. All that graphs are shown here figure 2.

Figure 2 Recorded (Time vs Amplitude) Voice Samples Frequencies Graphs

3.1.2) Second phase of Development: *Matching the current voice sample (Real Time) and send the data To the Parallel Prot for Command Execution*

In this step we compare the recorded voice template's frequency with the Current Voice Sample's frequency (real time speech). The basis of our comparison is the difference which we have calculated after recording of our command, here we compare the difference between the current voice sample frequency and the recorded sample's frequency and the execute the Functions corresponding to that Comparison, as shown in table 1 that Matched=1 for first command, every Time whenever the Current speech sample matched with recorded speech sample then a pop up window is generated Showing "Command Executed" and the color of the Stored Voice Sample panel is getting green from its base color(light gray) and blinks up to 3 mille seconds and also the Current Command Status Subplot is also gets blink green for the same Time and send binary code to Robotic Hand using parallel port(lpt1) communication which performs particular operation related

to that keyword's frequency. This process is very simple for key words (multilingual) recognition.

3.2) Development of Robotic Hand and Hardware for controlling the Robotic Hand

To setup the hand we chose a Simple Model of a Robotic Hand, we Implement the Three fingers of the Robotic hand, the parts used to setup a Robotic hand Shown in the Table 2 below

Sr. No.	Parts Used For Setup	
	Parts	Used For
1.	Aluminum	Designing the Hand base ,Palm base and Fingers
2.	Acrylic	Designing of palm
3.	Stepper Motor	Rotation of the Fingers(one for Each Finger)
4.	DC Motor(7 RPM)	Rotation of the Wrist(7 RPM for smooth Rotation)

Table 2 Parts Used for Setup of Robotic Hand

The given Table 2 Describes about the parts used for the development of the Robotic Hand. The figure 3 blow here gives the actual Setup of the Robotic Hand
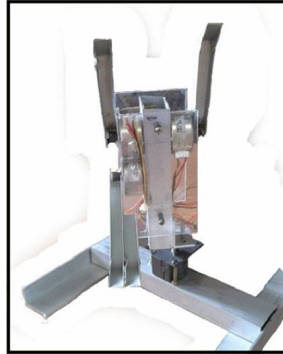
Figure 3 Robotic Hand

When Current command word given by user through microphone is recognized in Matlab® Binary code will be generated. This Binary code given to this Robotic Hand through the Parallel Port (lpt1) which is attached to the Hand and will perform particular operation related to that key word.

4. Results

As above in our Software development phase we have already discussed about the comparison of the current voice tamplets frequency with the recorded samples frequency according to the given comparison the corresponding functions are executed to control the robotic hand. I calculate the difference and recognize the keyword (Hindi or Punjabi) with the Following Commands,before that we have to take the current voice sample (real time) as our input because only after that we can apply the difference and recognizing operations for that purpose we used the following Command Sets

- *current_voice = wavrecord(t,fs)*
  Here this Command store the Current voice sample
- *[current_voice_current_voice_y]=find(current_voice >0.8);*
  Here this command assure that the value of the current sample should be above 0.8 magnitude that means it will take only value above this by this we will remove the noise level up to 0.8 magnitude and discarded the all below samples

- **diff_current_voice = max(current_voice_x) – min(current_voice_x);**

Here this command finds the difference between maximum and minimum value of current speech sample. Which is used further for the comparison purpose

- **if (diff_current_voice > limit0 && diff_current_voice < limit1)**
matched =1;

Here the command is used to calculate the difference between the current voice template's frequency and the recorded voice sample's frequency whose difference is stored in between the limit0 and limit1 for the first command, after successful comparison the associated Function matched =1 is executed and Corresponding to that the open hand command is executed.

The resultant of the Successful Comparison is as follows, it plot the graph for every successful match as shown in below figure 4
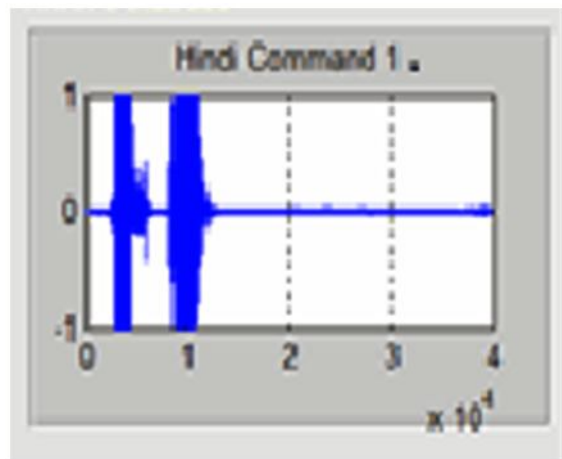

Figure 4 Graph of "हाथ खुले" Word Matched

Six key words (voice sample) three Hindi and three Panjabi here are used in this project to control the Robotic hand. User said "हाथ खुले" and this key word recognized by comparing with the recorded "हाथ खुले" key

word as shown in figure 4 above, as the recognizing procedure continue it is same for all the recorded Voice samples with the current voice samples the subplots for all the other voice samples.

The working of robotic hand will completely depend upon the Commands which are executed after the successful matching of the Current voice sample's frequency with the recorded voice sample's frequency.

5. Limitations of the work

This system is a Prototype system so that a very limited number of commands are recorded and implemented.In that way the accuracy of the system is almost 100%, consider the environment of room or lab is noisefree. In our future work we will try for more advanced and noise adaptable version of the system with with more commands.

6. Conclusion and Future Work

A fairly simple voice signal frequency system has been constructed using the speech data provided by the user through the mike using the power of the Matlab®. Here implemented procedure is very simple and yet use the flexible behavior of Matlab® to complete the desired tasks. The results obtained here demonstrate that the proposed simple algorithm is functional and it can be used in the many of the application for controlling devices specially in Robotics.Here our main goal was to develop a convenient algorithm for the Keyword Recognition which we have achieved with this project, the main achievement which is achieved here that we have taken the Hindi and Punjbai keywords for reorganization which used the Power of ASR,which was still be used for the English keyword reorganization, here our main effort will be now to develop the more convenient and noise effective algorithms for the future use in Humanoids with the use of Fuzzy Logic.

6. Acknowledgement

Firstly I wish to thank my guide, Asst. Professor Gianetan Singh sekhon, In-Charge, Computer Engineering Section, YCOE, Punjabi University, Guru

Kashi Campus, Talwandi Sabo. He has been supportive since the days I began working on the Robotics and ASR Ever since, he has supported me not only by providing a research assistantship over almost one year, but also academically and emotionally. Thanks to him I had the opportunity to build Robotic hand control system using Matlab. He helped me come up with the Present topic and guided me over almost a year of development. And during the most difficult times when writing this Paper, he gave me the moral support and the freedom I needed to move on. I would like to thank my parents, for supporting me throughout my life and above all I thank 'GOD' for making this venture possible.

7. References

[1] Abhishek Thakur et al.,(IJAEST)" International Journel of   Advance Engineering Sciences and Technologies" Vol No. 8, Issue No. 1, 100 – 106,2011

[2] Jamel Price and Ali Eydgahi, "Design of Matlab®-Based Automatic Speaker Recognition Systems," 9th International Conference on Engineering Education San Juan, PR,   July 23 – 28, 2006

[3] E. Darren Ellis, "Design of a Speaker Recognition Code using Matlab". Department of Computer and Electrical Engineering – University of Thennessee, Knoxville Tennessee 37996 09 May 2001

[4] Kerstin Dautenhahn. The aisb'05 convention-social intelligence and interaction in animal, robots and agents. In AISB'05:Social Intelligence and Interaction in Animal, Robots and Agents-SSAISB 2005 Convention, pages i–iii, Hatfield,UK, April 2005.

[5] Shafkat Kibria, "Speech Recognition for Robotic Control," Ume°a University Department of Computing Science, December 18, 2005.

* * * * *

Answers:

Q. 1. Poor English, Need lot of revision

Q2. This is expected to be the idea of the author and how it can be referenced from other's work. (Near Devices[1])

Ans. The current work is based upon the applications of the signal processing tool box of MATLAB®. The hardware setup is original, in home development. During the internet search the paper refer [1] was found which was similar in term of concept.

Q 3. On what factors the time frame duration and frequency depends. How authors jumps to these fugures i.e. 4000 to 8000 Hz?

ANS. A standard measure of frequency is Hertz, abbreviated Hz. It means "cycles per second." The sampling rate, sample rate, or sampling frequency (f_s) defines the number of samples per unit of time (usually seconds) taken from a continuous signal to make a discrete signal. These results have been achieved through the trial & error. 8000 HZ is the default sample rate for 1 second to store the voice sample in MATLAB®. There of 5 sec has been taken as it is time duration in which a normal user can save his command.

Q. 4. What is the need of using start and stop keywords in English? Can't we use any Punjabi or Hindi word for the same?

ANS. These line have been deleted since the procedure is too basic to explain here.

Q. 5. Contradictory statements. Author jumps from several phases to two phases.

ANS. Thank you, I will do necessary corrections.

6. What is the effect of difference of freq. of voice templates on the system performance and how author decide this diff to be 2000?

ANS. Lines removed as the given process is Part of MATLAB® Implementation.

Q 7. Misplaced Table.

ANS.

Q 8. What does 0.6 signifies?

ANS. 0.6 is threshold value of the voice sample frequency amplitude. Here in speech processing toolbox the maximum value of this amplitude is 1. so here we have taken this value so that the voice signal containing under this value will be discarded automatically when we recorded the voice sample.(e.g whenever we are trying to record the voice sample in Matlab environment with the help of mike the noises at room level are also recorded by the mike along with the voice sample so to make the voice sample noise free we set this threshold value so that the voice frequency below this will automatically discarded. please refer the below image for more
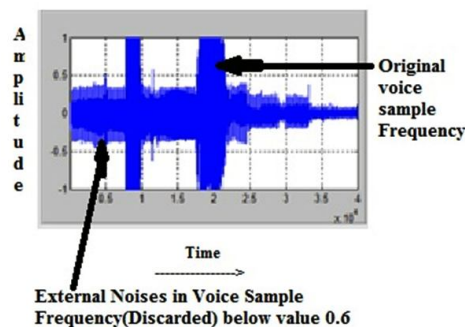
Voice Sample Recorded For the Keyword "Hath Kholo"

Q.9. From where this figure 0.6 arrives?

ANS. It is an observation that has been made by running the system in MATLAB®

Q.10. Why samples <0.6 are discarded?

ANS. In order to eliminate noise signals whose values is generally < 0.6.

Q.11. What do you mean by value of speech sample? Are you pointing to freq or some other parameter?

ANS. Here speech sample means the voice we have recorded in our data base for matching or the voice we are recording while the project is running (current voice).

Q.12. What does these graphs shows? Justify their presence here..

ANS. The graphs here are plotted in time and amplitude axis shows the frequency of each voice sample recorded.

Q.13. Earlier Author used 0.6 now 0.8 why?

ANS. 0.6 for the less noisy environment and 0.8 is for the more noisy environment. In general their values can be adjusted as per the noise level of the environment.

Q.14. From where the values of limit0 and limit1 arrives?

ANS. limit0 and limit1 are the functions which are taken during the implementation in which the difference between two voice frequencies are stored and that difference is used for the purpose of matching.

Q.15. How diff_current_voice and Diff_hindi_word relates to each other.?

ANS. Diff_hindi_word is the difference between the max and min values of the amplitude of the hindi voice sample frequency(saved in the database) and the Diff_current_voice is the difference between the max and min values of the amplitude of the current voice sample frequency(taken while the project is running).

Q.16 The steps discussed in results are actually part of implementation. It is better if you discuss its accuracy of system in results section and any other relevant parameter on which system can be checked should also discuss problematic area.

ANS. The results of the project totally depends upon the accurate execution of the given commands, to that matter I have discuses these steps in the result section.   This system is a Prototype system so that a very limited number of commands are recorded and implemented. In that way the accuracy of the system is almost 100%, consider the environment of room or lab is noise free.

Q.17 Incomplete reference among references.

ANS. All incomplete references are corrected now.

Q.18 Duplicate references.

Ans. All duplicate references are corrected now.

* * * * *