Concurrent Voice Transmission with Customized Grammar Rules based on Locale

Komal Chauhan
Department of Computer Science & Engineering
Amity University, Noida, India

Kamal Kant
Department of Computer Science & Engineering
Amity University, Noida, India

ABSTRACT
Being visually impaired or a computer illiterate have always being a barrier for people to use computer to perform their tasks in an easy, efficient and quick way. However, mobile is not only a necessity of everyone today but is convenient to be operated by anyone and everyone. This paper discusses a system wherein the user will be able to type text on computer by providing a voice input through his mobile phone.

General Terms
Speech Recognition

Keywords
Speech Recognition System, MFCC, HMM, N-Gram Dataset.

1. INTRODUCTION
Irrespective of age, gender, educational background or physical impairment (excluding being dumb), speech is the primary mode of communication used by human beings to express themselves to others. It is this human speech which forms the basis of one of the hottest field of modern science i.e. Speech Recognition. The research work presented in this paper uses a mobile phone to provide speech input to a speech recognition system which in turn gets printed as text on the computer screen This paper is divided into two major parts, in the first part an introductory overview of speech recognition system is provided along with a glimpse of some of the recent researches in the field so that the reader gets a background of this field and the second part discuses the research work under consideration in this paper.

2. SPEECH RECOGNITION SYSTEM

2.1 Definition
Popularly known as Automatic Speech Recognition (ASR), speech recognition is a sub-category of pattern recognition wherein the speech input is first understood by the computer to perform the user desired task using it. Its preliminary task is not only to respond instantly but also to work effectively in noise or complete silence environment and that too on heterogeneous inputs.

2.2 Types of Speech [1] [8]
Different ASR systems accept speech input in different forms.

- Isolated Words: Isolated i.e. single utterances are fed as input to the recognizer but choosing word boundaries affects result obtained.
- Connected Words: Separate utterances together with minimum pause are input requirement of this system.
- Continuous Speech: A dictation by computer to the speaker, it is the most difficult recognizers to create.
- Spontaneous Speech: Speaker's natural speech acts as the input for the system.

2.3 Stages of Speech Recognition Technique [2] [3]
1 Analysis: Vocal tract, excitation state characteristics and behavior characteristic of the speaker are identified.

2 Feature Extraction: Spectral features along with excitation source are identified. This stage is further divided into two steps. The first is the training step, shown in Fig. 1 below, wherein the system is familiarized with the speaker's voice characteristics and these act as reference models.

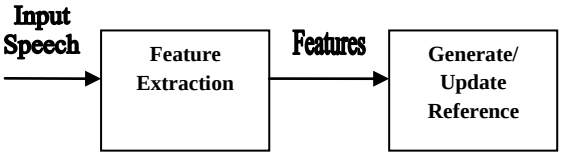


Fig 1: Training [7]

The second step is testing, shown in Fig. 2, where unknown utterances are matched with the reference model to find their best possible match.

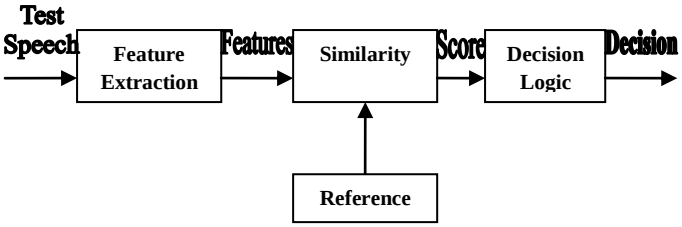


Fig 2: Testing [7]

3 Modeling Technique: Here speaker models are generated which are categorized into speaker identification (where input speech signal helps to identify the speaker) and speaker recognition. Speaker recognition is further

divided into speaker dependent (where the extracted characteristics identify speaker) and speaker independent (where only the content matters and not who is speaking it). The various modeling techniques available are:

- *Acoustic-Phonetic Approach:* Based on acoustic properties labels are allotted to speech sounds.
- *Pattern Recognition Approach:* First the system is trained with utterances which act as reference patterns and then unknown utterances are compared to these references to know their identity.
- *Template Based Approach:* In this there is a candidate word dictionary which acts as templates. Input unknown utterances are matched with these templates and the one that matches the best is selected. However, the production and storage of template per word is an impractical task.
- *Dynamic Time Warping (DTR):* It measures the similarities between two utterances that vary in speed and are "warped" non-linearly in time dimension, on a frame-by-frame basis.
- *Knowledge Based Approach:* The knowledge of experts related to variation in speech is hand coded in the system, though it is not easy to obtain such knowledge and use it effectively.

4 Matching Technique:

- *Whole Word Matching:* There is a large storage requirement in this approach as all incoming speech signals are compared to these pre-recorded word templates. Results are obtained quite quickly.
- *Sub-word Matching:* Recognition is done with the help of phonemes. Though there is less storage requirement for templates but processing time is more.

3. HIDDEN MARKOV MODEL: SPEECH RECOGNITION TECHNIQUE

As discussed above training and testing are the two phases of speech recognition. Hidden Markov Model i.e. HMM is a successful and flexible technique used in speech recognition. In this research work also HMM is used so a brief overview of it is given in this section.

3.1 Introduction of HMM

An extension of Markov model, HMM [6] [31] is a model in which the intermediate states that are responsible to transform the initial state to final output state are hidden. The triplet which represents HMM is given as:

$$(A, B, \pi) \quad (1)$$

Where A is state-transition probability

B is observation/output probability

π is initial state

HMM is a probabilistic model wherein the model can go from one state to another within a time shift but it is only probabilistic.

3.2 Elements of HMM

- N number of states in the model - Though the principle of HMM is that it's states are hidden but still they have physical significance.
- M number of distinct observation symbols per state – It represent the physical output emitted by the model under consideration.
- S represents the individual state.

3.3 Steps in HMM

1. Evaluation: In this step the probability of a model to generate an observation sequence is judged so as to find the best model available. For the HMM model λ, the observation sequence O is given as :

$$P(O \mid \lambda) \quad (2)$$

2. Decoding: It is the process wherein the best state sequence, Q, is obtained for the observation sequence, O.
3. Training (Learning): The most tedious step of HMM where the model parameters (A, B, π) are adjusted to maximize the observation sequence probability.

3.4 Basic Problems of HMM

Problem 1: For the given observation sequence O and a model $\lambda = (A, B, \pi)$, how to efficiently find the best model i.e. $P(O \mid \lambda)$.

Problem 2: For the given observation sequence O and a model λ, how to choose the state sequence Q which best explains the observation.

Problem 3: How the model parameters (A, B, π) should be adjusted to maximize $P(O \mid \lambda)$.

4. MEL FREQUENCY CEPSTRUM COEFFICIENT (MFCC): FEATURE EXTRACTION TECHNIQUE

MFCC is the feature extraction technique employed in this research work which not only extracts but also selects the parametric representation which is best for acoustic signal. The steps involved in MFCC calculation are:

1. *Mel Frequency Wrapping*: The pitch under consideration is measured using a 'mel' scale as human speech does not follow linear scale for measuring frequency of speech signal. 'Mel' scale spaces linearly below 1000 Hz and logarithmically above 1000 Hz. Mathematically it is formulated as:

$$Mel(f) = 2595 * \log 10(1 + f / 700) \quad (3)$$

2. *Cepstrum*: The log mel spectrum obtained from above step, are converted to time from real numbers using discrete cosine transform, which represent local spectral properties of the speech signal.

$$C_n = \sum_{k=1}^{k} (\log S_k) \cos \{n(k-(1/2)* \pi /k)\} \quad (4)$$

5. PERFORMANCE OF SYSTEMS

Rate of accuracy and speed are the two parameters that measure the effectiveness of any speech recognition system [7]. Word Error Rate (WER) measures accuracy and is given in equation (5):

$$WER = (S + D + I) / N \quad (5)$$

Where S are the number of words that are substituted

D are the number of words that are deleted

I are the number of words that are inserted

N are the total number of words under consideration

Speed is measured in terms of Real Time Factor (RTF) which is given in equation (6):

$$RTF = P / I \quad (6)$$

Where P is the time that will be taken to process input of I duration

Speech recognizer is widely used in the field of voice authentication where the recognizer on the basis of input speech signal judges the authenticity of the speaker. The speaker should neither be authenticated by the recognizer when he should not be nor should the speaker be unauthenticated when he should i.e. a balance needs to be maintained between false acceptance rate (FAR) and false rejection rate (FRR). On plotting a graph between FAR and FRR, as shown in Fig. 3, the point of intersection is known as crossover error rate (CER) which must be low for a better system performance.

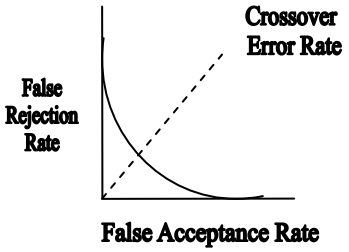


Fig 3: FAR, FER, CER [4]

5.1 Analysis of Existing Systems

The table 1 below brings into picture some of the existing systems whose accuracy varies on the basis of which feature extraction technique they choose out of LPC (Linear Predictive Coding), MFCC (Mel Frequency Cepstral Coefficient) and PLP (Perceptual Linear Prediction) and which recognition technique they choose out of HMM (Hidden Markov Model), GA (Genetic Algorithm) and VQ (Vector Quantization).

Table 1. Analysis of Existing Speech Recognition Systems

Research Work Name	Feature Extraction Technique	Recognition Technique	Accuracy
Alaigal-A Tamil Speech Recognition [9]	PLP	HMM	70% - 80%
Hindi Speech Recognition System Using HTK [11]	MFCC	HMM	Word-accuracy and word-error rate of the system are 94.63% and 5.37% respectively
Speech Emotion Recognition System based on Integrating Feature and Improved HMM [12]	MFCC	HMM + GA	More than 77%
Continuous Speech Recognition System for Tamil Using Monophone-based Hidden Markov Model [15]	MFCC	HMM	92% accuracy in word level and 81% accuracy in sentence level
Automatic Speech Recognition for Bangla Digits [17]	MFCC	HMM	More than 95% for digits (0-5) and less than 90% for digits (6-9)
Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK) [19]	MFCC	HMM	97.99%
English Digits Speech Recognition System Based on Hidden Markov Models [27]	MFCC	HMM	56.25% - 72.5%
Human Computer Interaction Using Isolated Words Speech Recognition Technology	MFCC	VQ	88%

[30]			
Segment-Based Stochastic Modelings for Speech Recognition [8]	LPC	HMM + VQ	62% - 96%
Automatic Speech Recognition:Human Computer Interface for Kinyarwanda Language [42]	MFCC	HMM	92%

6. PROPOSED SYSTEM

Initially the input signal is transformed from analog to digital form and then it is divided into frames each having their individual frequency. The spectral features extracted from the frames help in identifying the phones of the speech sample which are nothing but sounds that distinguish two words. MFCC is applied to the phones which identify unique discrete acoustic phone of each individual input speech sample. Then HMM is applied to each phone which identifies the likelihood of occurrence of a word W within the given acoustic observation, P(W/A). This can be represented using Baye's rule format:

$$P(W/A) = P(A/W)P(W)/P(A) \quad (7)$$

Where P(A/W) represents the acoustic model

P(W) represents the language model giving the probability of sequence of words

P(W) is calculated using the Microsoft N-Gram dataset which has the ability to predict the next phone, letter or word in a given sequence of input. The N-Gram dataset is being used here as a data source since it is a huge repository of data with data being collected from world web pages and internet documents, which have data ranging from proper nouns, domain specific terms, special expressions, technical words, acronyms and terminologies; covering an ample number of words of the language. With the help of N-Gram dataset, for the proposed system vocabulary dictionary and sentences are available which act as a huge repository for both training as well as testing phase of the proposed system. An algorithm called soundX is then created. This algorithm finds the best possible match from all the possible patterns returned by N-Gram dataset. For eg: If the sentence being provided as input speech is: "Can I kiss the baby!", then the N-Gram dataset might generate possible available matches "kill" and "kiss". However, the soundX algorithm will consider the pattern with highest occurrence frequency and select it i.e. in this case "kiss" will fit into the above sentence. In other words, from all the possible matches returned by the N-Gram dataset, soundX selects the one that fits the best. This sentence thus selected with the help of N-Gram dataset and soundX algorithm is printed as text on screen. The soundX algorithm is splitted into three parts major parts as shown below:

```
Function DetectError (In)

// In this voice input is fed to the ASR system

{

// split the text received as input by ASR and return word tokens
WSplit(In,“”)
for(i <- 0 to i <- N) // detect all word tokens
{
// search for W[i] in Microsoft N-Gram dataset where W denotes a particular word
R <- Search(N-Gram Dataset , W[i])
if(R == true) // i.e. if W[i] is found in Microsoft dataset
i <-  i+1 // go to the next word token W[i+1]
else
//W[i] is misspelled and thus a correction is required so go to the candidate corrections generation algorithm
AlternateCandidates(W[i])
}
}
Function AlternateCandidates (word)
{
// create 2-gram character sequences and put them in an array named a
a <-Split2Grams(word)
for(i <- 0 to i <- N) // for all 2-gram sequences
{
// look for unigrams having a[i] as substring
L[i] <- Substring(N-Gram Dataset, a[i])
i <- i+1
}
// select the top 4 unigrams sharing 2-gram character sequences with the incorrect word
Candidates <- commonUnigrams(L)
// go to the error correction algorithm
CorrectError(candidates)
}
Function CorrectError (candidates)
{
for(i <-0 to i <- N)
// process all alternate candidates returned
{
// concatenate together the ith candidate with the preceding words
// A is a global array containing the original ASR output text
O <- Concatenate(A[j-4] , A[j-3] , A[j-2] , A[j-1] , candidates[i] )
// find O in  N-Gram dataset and return its frequency
frequency[i]  <-  Search(N-Gram Dataset , O)
i <- i+1
}
f <- MaxFrequency(frequency)
// return the index f of the candidate whose O has highest frequency
// return the correction for the ASR error
GENERATE_TEXT candidates[f]
}
```

The above soundX algorithm carries out the required filtration of patterns returned by N-Gram dataset to find the one that best matches the desired requirement. This work has reduced human effort to a great extent as human can speak more quickly than typing via a keyboard. Also while a speaker provide voice input he can utilize his hands for any other work. Moreover the system can also prove to be a blessing for

visually challenged people or the ones which don't have hands.

7. EXPERIMENTAL SETUP

Fig. 4 below demonstrates the experimental setup for the proposed work.

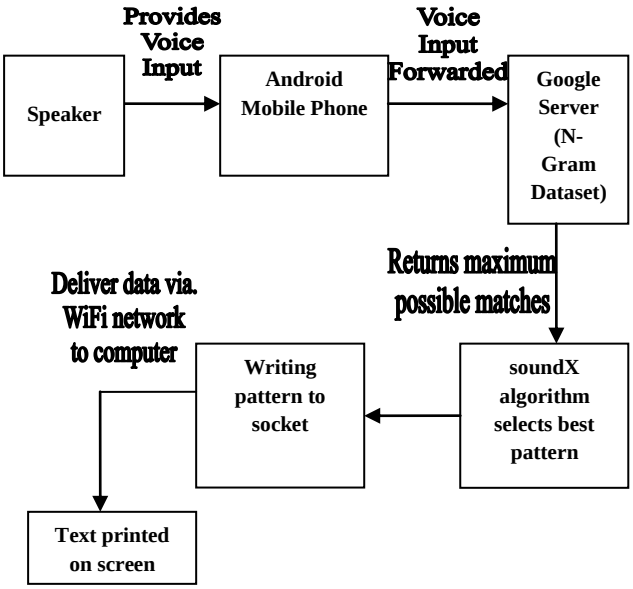
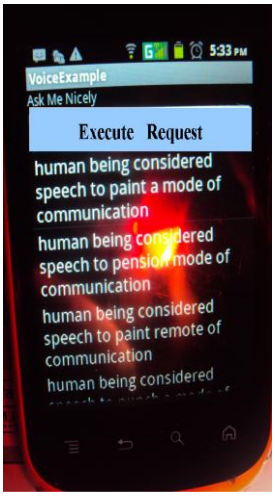


Fig 4: Experimental setup for the proposed system

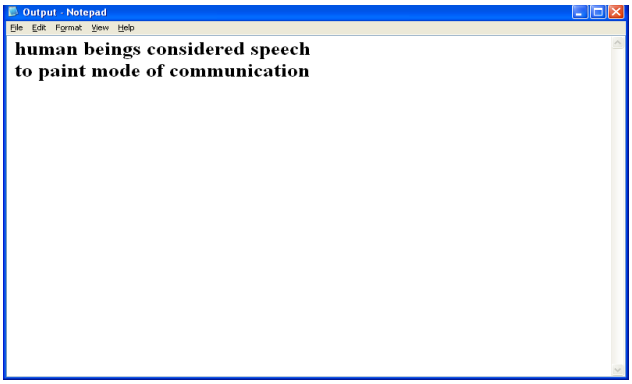
The objective of the proposed system is to print text on computer screen by providing voice input and not by using traditional means like keyboard or mouse but by providing input through android mobile phone. From the speech input, features are extracted using MFCC and the acoustic characteristics are recognized using HMM. MFCC and HMM are basically used to obtain the speech signal in it's maximum possible pure form i.e. devoid of noise or any other impurities that are fed along with speech input. This modified signal is fed to Google server which returns the maximum possible matches for the input speech signal by making use of N-Gram dataset. Then soundX algorithm is designed which makes use of NLP (Natural Language Processing). soundX selects the best pattern from the lot obtained from Google server. To send this pattern to the computer, so that it can be printed on screen, socket programming is used which uses WiFi network as a medium of transmission. This is how the desired task is performed.

8. RESULTS

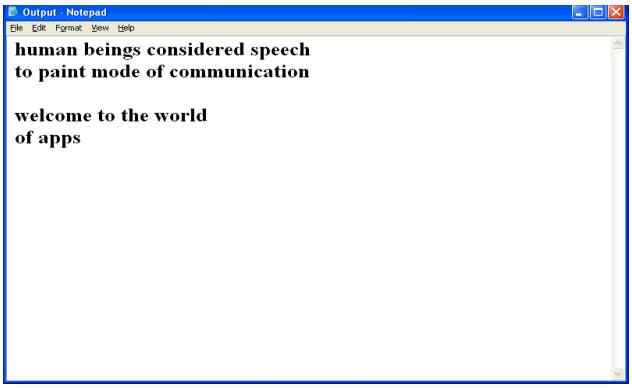
The following interfaces depict how the system is working. Initially the user opens app called VoiceExample in his android phone and presses the "Execute Request" button. User then provides some voice input. For e.g. here the user provided the voice input "Human beings consider speech as the primary mode of communication". This voice input goes to the refinement process using MFCC and HMM and then goes to Google server which makes use of N-Gram dataset and returns the best possible matches for the input provided, as shown below:



From these matches returned, the soundX algorithm designed during this proposed system, selects the best possible match and sends it to the computer screen to be typed as shown below.



Now if user again provides some input say "Welcome to the world of apps", then after going through all the processing as before, this new input gets appended after the initial input as shown below. The notepad file gets refreshed on it's own, the user does not need to do anything to make updation in it.



The above interfaces are depicting the accomplishment of the objective with which this proposed system was started i.e. text being typed on screen by providing voice input is being accomplished.

As discussed in the paper above RTF and WER are the two performance parameters of a speech recognizer. For this system RTF is calculated as:

RTF = 4.6/5 = 92%

where 4.6 seconds P processing time is taken for an input of duration I of 5 seconds.

On the basis of the interfaces shown above, WER is calculated for the input speech: "human beings consider speech as the primary mode of communication whose output that is returned by the system is: "human beings considered speech to paint mode of communication". As it is evident from the interfaces, for N = 10 total words spoken by speaker; S = 3 substitutions are made as 'consider' is replaced by 'considered', 'as' is replaced by 'to' and 'primary' is replaced by 'paint'; I = 0 as no insertions are made; D = 1 as 'the' is deleted from the N-Gram dataset returned output. Therefore, WER is:

WER = (3+0+1)/10 = 40%

Experimentally it is believed the lower is the WER, the more accurate is the speech recognizer. However, WER variation depends on factors such as age, accent of speaker and number of words fed as input to the recognizer. The following graphs depict variation in WER on basis of these factors. The graph in Fig.5 is plotted between WER and age, with age varying from 20, 30, 40, 50, 60 years. The graph in Fig.6 is plotted between WER and no. of words fed as input per processing i.e. starting from 1 word for 1 processing to words, 3 words, 4 words and 5 words together given for processing.. The graph in Fig. 7 is plotted between WER and English accent as per mother tongue which ranges from Punjabi accent English, Tamil accent English, British English accent, Haryanvi accent English, Urdu accent English.

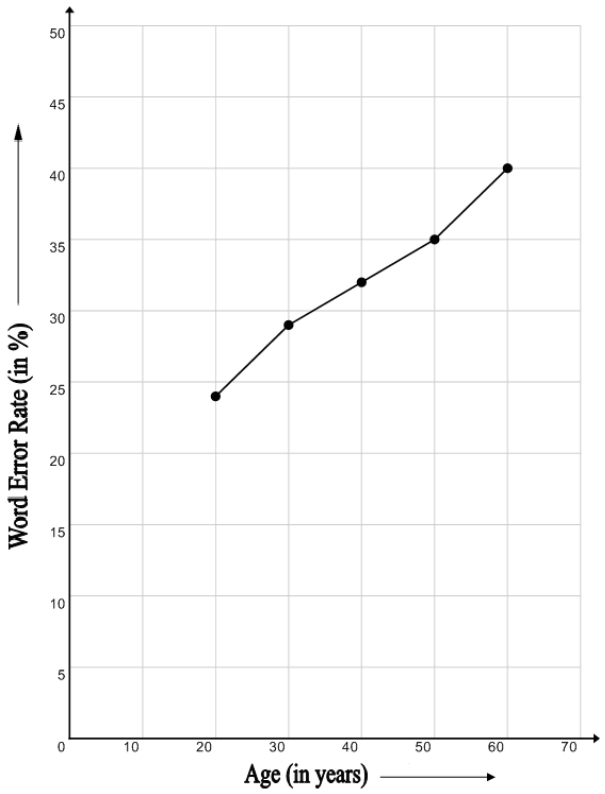


Fig 5: Graph plotted between WER and Age

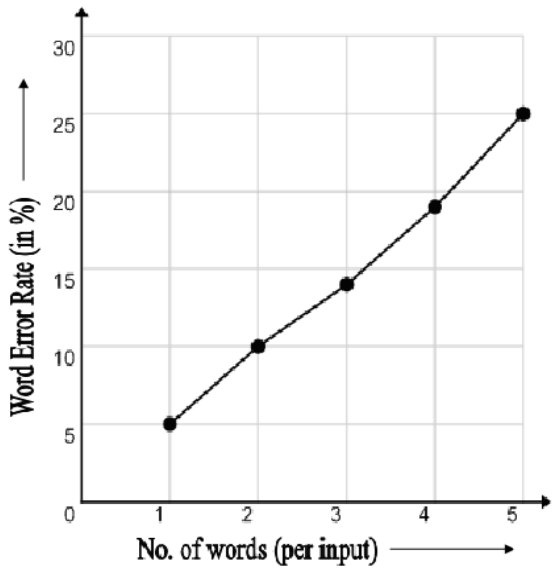


Fig. 6 Graph plotted between WER and No. of words per input

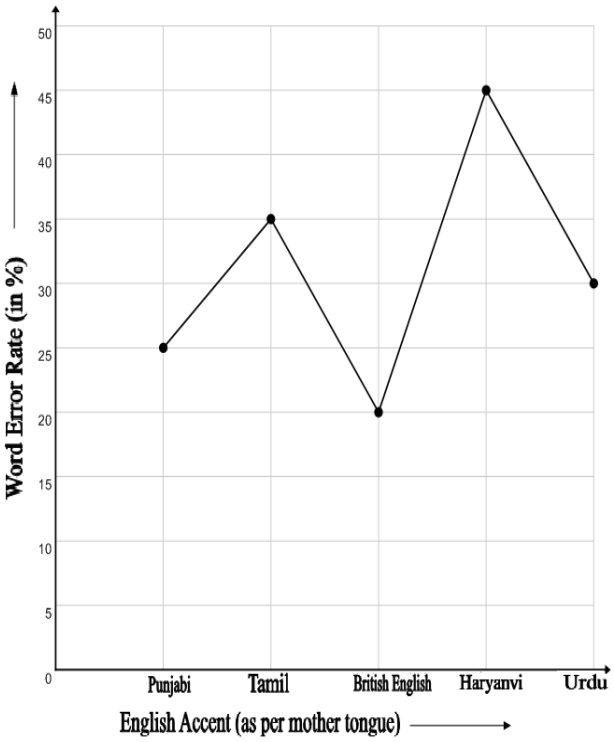


Fig. 7 Graph plotted between WER and English Accent (as per mother tongue)

9. CONCLUSION AND FUTURE WORK

More than 50 years is the legacy of speech recognition system. The ASR systems prove to be useful not only for blind people but also let able people to do some other work as their hands and eyes are free to indulge in other activities. Moreover, these ASR systems can be developed in and are existent in native languages like English, Tamil, Punjabi, Hindi, Chinese, etc; which breaks the obstacle of the person being educationally underprivileged to operate the computer. Speaker independent continuous speech recognition systems with large vocabulary are in-demand which can be fulfilled by using the feature extraction technique MFCC with the recognition technique HMM which help in creating extremely powerful systems that offer good speech recognition results.

For future work, a new system can be designed on similar grounds where not only text is being printed on screen but most of the computer functioning can be handled by providing voice input through mobile phone. This can range from typing URL in browser's address bar to computer's control functions like: Ctrl+C, Ctrl+X; to name a few to attaching files to e-mails. Moreover, currently this proposed system requires both the mobile and laptop to be in same network. But in future the entire system can be deployed in GPRS so that the mobile and laptop can coordinate from remote locations.

10. REFERENCES

[1] Pradeep Kumar Jaisal, Pankaj Kumar Mishra, "A Review of Speech Pattern Recognition Survey", International Journal of Computer Science and Technology, 2012.

[2] Santosh K.Gaikward, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications, 2010.

[3] "Speech Recognition Technology Choices", A Vocollect White Paper, 2010.

[4] Lisa Myers, "An Exploration of Voice Biometrics", SANS Institute Infosec, 2004.

[5] Rudan Bettelheim, David Steele, "Speech and Command Recognition", FreeScale White Paper, 2010.

[6] Bhupinder Singh, Neha Kapur, Puneet Kaur, "Speech Recognition with Hidden Markov Model : A Review", International Journal of Advanced Research in Computer Science and Software Engineering, 2012.

[7] Vibha Tiwari, "MFCC and it's Applications in Speaker Recognition", International Journal on Emerging Technique, 2010.

[8] Vimala C., Dr. V.Radha, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal, 2012. A.P.

[9] Henry Charles G. Devaraj, "Alaigal- A Tamil Speech Recognition", Tamil Internet Singapore, 2004.

[10] Er. Jaspreet Kaur, Er. Nidhi, Ms. Rupinder Kaur, "Issues Involved in speech To Text Conversion", International Journal of Computional Engineering Research, 2012.

[11] Kuldeep Kumar, R.K.Aggarwal, "Hindi Speech Recognition System using HTK", International Journal of Computing and Business Research, 2011.

[12] Han Zhiyan, Lun Shxian, Wang Jian, "Speech Emotion Recognition System based on Integrating Feature and Improved HMM", The 2nd International Conference on Computer Application and System Modeling, 2012.

[13] Ore A. Soluade, "A Comparative Analysis of Speech Recognition Platforms", Communications of the IIMA, 2009.

[14] M. Chandrashekhar, M.Ponnavaikko, "Tamil Speech Recognition: A Complete Model", Elctronic Journal Technical Acoustics, 2008.

[15] Radha. V, Vimala. C, Krishnaveni. M, "Continuous Speech Recognition System for Tamil Language Using Monophone-based Hidden Markov Model", CCSEIT-12, ACM, 2012.

[16] Zhao Lishuang, Han Zhiyan, "Speech Recognition System Based on Integrating Feature and HMM", IEEE DOI 10.1109/ICTMA.2010.298, 2010.

[17] Ghulam Muhammad, Yousef A. Alotaibi, Mohammad Nurul Huda, "Automatic Speech Recognition for Bangla Digits", IEEE 978-1-4244-6284-1109/ICCIT.2009, 2009.

[18] Marek Bohac, "Performance Comparison of Several Techniques to Detect Words in Audio Streams and Audio Scene", IEEE 54th International Symposium ELMAR-2012, 2012.

[19] Bassam A.Q. Al-Qatab, Raja N. Ainon, "Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK)", IEEE 978-1-4244-6716-711 0, 2010.

[20] Sandipan Mandal, Biswajit Das, Pabitra Mitra, "Shruti-II: A Vernacular Speech Recognition System in Bengali and an Application for Visually Impaired Community", IEEE 978-1-4244-5974-2/10, 2010.

[21] Xiaofeng Wu, Ryohei Nakatsu, "Vision-aided Speech Recognition System for a Small Four-Legged Robot", IEEE 978-1-4244-5858-5/10, 2010.

[22] Sanjana Primorac, Mladen Russo, "Android Application for Sending SMS messages with Speech Recognition Interface", MIPRO, 2012.

[23] Yanan Jian, Jianshe Jin, "An Interactive Interface between Human & Computer based on Pattern & Speech Recognition", International Conference on Systems & Informatics, IEEE 978-1-4673-0199-2122, 2012.

[24] Huda Sarfaraz, Sarmad Hussain, Riffat Bokhari, Agha Ali Raza, Inam Ullah, Zahid Sarfaraz, Sophia Pervez, Asad Mustafa, Iqra Javed, Rahila Praveen, "Large Vocabulary Continuous Speech Recognition for Urdu", ACM 978 1-4503-034202/10/12, 2010.

[25] Matus Pleva, Stanislav Ondas, Jozef Juhar, Anton Cizmar, Jan Papaj, Lubomir Dobos, "Speech & Mobile Technologies for Cognitive Communication & Information Systems", IEEE Proceedings, 2nd International Conference on Cognitive InfoCommunications (CogInfoCom), 2011.

[26] Oscal T.C. Chen, Jhen Jhan Gu, Ping-Tsung Lu and Jia-You Ke, "Emotion Inspired Age and Gender Recognition Systems", IEEE 978-1-4673-2527-1/12, 2012.

[27] Ahmad A.M. Abushariah, Teddy S. Gunawan, Mohammad A.M. Abushariah, Othman O. Khalifa, "English Digits Speech Recognition System Based on

Hidden Markov Models", International Conference on Computer & Communication Engineering, 978-1-4244-

6235-3/IEEE, 2010.

[28] A. Mukhopadhayay, S. Chakraborty, M. Choudhary, A. Lahiri, S. Dey, A.Basu, "Shruti: An Embedded text-to speech system for Indian Languages", IEEE Proceeding-Software Vol.153, No.2, April 2006.

[29] M.U. Akram, M. Arif, "Design of an Urdu Speech Recognizer based upon acoustic phonetic modeling approach", 0-7803-8680-9/04, IEEE, 2004.

[30] Mohammad A.M. Abu Shariah, Raja N. Ainon, Roziati Zainuddin, Othmn O. Khalifa, "Human Computer Interaction Using Isolated Words Speech Recognition Technology", International Conference on Intelligent & Advanced Systems, 1-4244-1355-9/07, IEEE, 2007.

[31] Mohd Zaizu Ilyas, Salina Abdul Samad, Aini Hussain, Khairul Anuar Ishak, "Speaker Verification using Vector Quantization and Hidden Markov Model", The 5th Student Conference on Research and Development, 1-4244-1470-9/07, IEEE, 2007.

[32] Che Yong Yeo, S.A.R. Al-Haddad, Chee Kyun Ng, "Animal Voice Recognition for Identification (ID) Detection System", IEEE 7th International Colloquium on Signal Processing & it's Applications, 2011.

[33] Charl van Heerden, Etienne Barnard, Michael Feld, Christian Miller, "Combining Regression & Classification Methods for Improving Automatic Speaker Age Recognition", ICASSP 2010, IEEE, 2010.

[34] Michael Grimm, Kristian Kroschel, Shrikanth Narayanana, "The Vera Am Mittag German Audio

Visual Emotional Speech Database", ICME 2008,

IEEE, 2008.

[35] Charles Corfield, "Demystifying Speech Recognition", nVoq White Paper.

[36] Z.Hachkar, B.Mounir, A.Farchi, J.El Abbadi, "Comparison of MFCC and PLP Parameterization", Canadian Journal on Artificial Intelligence, Machine Learning & Pattern Recognition, 2011.

[37] Lidia Mangu, Mukund Padmanabhan, "Error Corrective Mechanisms for Speech Recognition", IEEE 0-7803-7041-4/01, 2001.

[38] S.Basu, C.Neti, N.Rajput, A.Senior, L.Subramaniam, A.Verma, "Audio-Visual Large Vocabulary Continuous Speech Recognition in Broadcast Domain", IEEE 0-7803-5610-1/99, 1999.

[39] Michael Feld, Etienne Barnard, Charl van Heerden, Christian Muller, "Multilingual Speaker Age Recognition: Recognition Analyses on the Lwazi Corpus", IEEE 978-1-4244-5480-8/09, ASRU 2009.

[40] F. Reena Sharma, S. Geetanjali Wason, "Speech Recognition & Synthesis Tool: Assistive Technology for Physically Disabled Persons", International Journal of Computer Science & Telecommunications [Volume 3, Issue 4, April 2012].

[41] M. Jackson, "Automatic Speech Recognition: Human Computer Interface for Kinyarwanda Language", Master Thesis, Faculty of Computing & Information Technology, Makerere University, 2005.

[42] Nelson Morgan, "Deep and Wide: Multiple Layers in Automatic Speech Recognition", IEEE Transactions on Audio, Speech & Language Processing, Vol. 20, No. 1, January 2012.