ISSN 2320 - 2602
Volume 2, No.12, December 2013
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2013/ijacst012122013.pdf

Speech Recognition in context of predefined words, phrases and sentences stored in a database and its analysis, designing, development and implementation in an application

Nadeem Ahmed Kanasro[1], Habibullah U.Abbasi[2], Abdul Ghafoor Memon[3], Mujeeb-U-Rehman Maree[4], Kamran Taj Pathan[5]

nadeemahmedkanasro@yahoo.com, habibullah.abbasi@gmail.com, ghafoor@usindh.edu.pk, mrmaree@yahoo.co.uk, kameetaj@gmail.com

Teaching Assistant, IMCS, University of Sindh, Jamshoro, Sindh (Pakistan)[1]
Assistant Professor, CFES, University of Sindh, Jamshoro, Sindh (Pakistan)[2]
Associate Professor, IMCS, University of Sindh, Jamshoro, Sindh (Pakistan)[3]
Assistant Professor, IMCS, University of Sindh, Jamshoro, Sindh (Pakistan)[4]
Assistant Professor, IICT, University of Sindh, Jamshoro, Sindh (Pakistan)[5]

ABSTRACT

This paper presents implementation and use of our designed Language Models and Grammar in Speech Recognition Engine and its analysis, designing, development and implementation in an application. This research is concerned with 'Speech Recognition Application' named 'Text Editor Through Voice', which is operated as 'Speech Recognition (Speaker Independent) System', based on 'Speech Recognition Technology'. The approach is based on experiencing the praxis using 'Hidden Markov Model' and application is designed in Visual Basic 6.0 using 'Visual Programming' and 'Object Oriented Programming' methods. In 'Text Editor Through Voice' the use of Speech Recognition engine translates spoken input before finding the specified words, phrases, and sentences stored in database. After finding and matching recognized input from database it puts that in document area of text editor just like typing on keyboard and pressing a key on the phone keypad, in this application microphone to be able to do this. For example one might say a word like "Hello", to which application replies by inserting said word in the document area. Furthermore, we show you list of words and phrases in tables with figures that are successfully implemented and executed in our developed application.

Keywords: Markov Model, Neural Network, Language Model & Grammar, Speech Recognition Engine, Dynamic Time Warping and Graphical User Interface (GUI)

1. INTRODUCTION

The designing and developing a computer application for a machine that mimics person activities, mostly the ability of talking naturally and responding appropriately to spoken language, has intrigued engineers and scientists for centuries. Since the 1930s, when Homer Dudley of Bell Laboratories projected a system model for speech study and synthesis [5], the problem of automatic speech recognition has been approached progressively, from a simple instrument that responds to a small set of sounds to a complicated system that responds to effortlessly spoken natural language and takes into description the varying information of the language in which the speech is produced. Based on major advances in statistical modeling of speech in the 1980s, automatic speech recognition systems today find extensive application in farm duties that require a human-machine interface.

Speech based applications are developed to perform different tasks such types of applications are given below;

- **Simple data entry:** These types of applications are used to enter numbers, characters, and phonemes. For example: entering a credit card number

- **Voice user interfaces:** These types of applications are used to make a call by (VCD) voice command device, these applications fall into different categories like
  - Voice activated dialing
  - Routing of Calls
- **Domestic appliance control:** These types of applications are used to control home appliances, for example: turn off tube lights, where particular words are spoken.
- **Preparation of structured documents:** These types of applications are used in medical science to create reports, for example: a radiology report.

- **Speech-to-text processing**: These types of applications are used to dictate, process spoken words, word processors or emails are examples of these applications.

Speech recognition is the transformation of verbal inputs known as words, phrases or sentences into content. It is also known as 'Speech to Text', 'Computer Speech Recognition' or 'Automatic Speech Recognition'. It is one kind of technology and was first introduced by AT&T Bell Laboratories in the year 1930s.

Some speech based software are used for dictation on desktop machines for example as users speak something into the microphone then these software types same spoken words, phrases or sentences on the screen.

Some speech recognition based systems use "trainings" where speakers read a chunk of text. These systems examine specific voice of the individual and use it to fine tune the detection of that person's speech, resulting in more correct transcription. Training based systems are called Speaker Dependent systems while non training based are called Speaker Independent systems.

The speech recognition process is performed by a software component known as the **speech recognition engine**. The primary function of the speech recognition engine is to process spoken input and translate it into text that an application understands.

Figure# 1 shows that Speech recognition engine requires two types of files to recognize speeches, which are defined below.

1. **Language Model** or **Grammar**
2. **Acoustic Model**

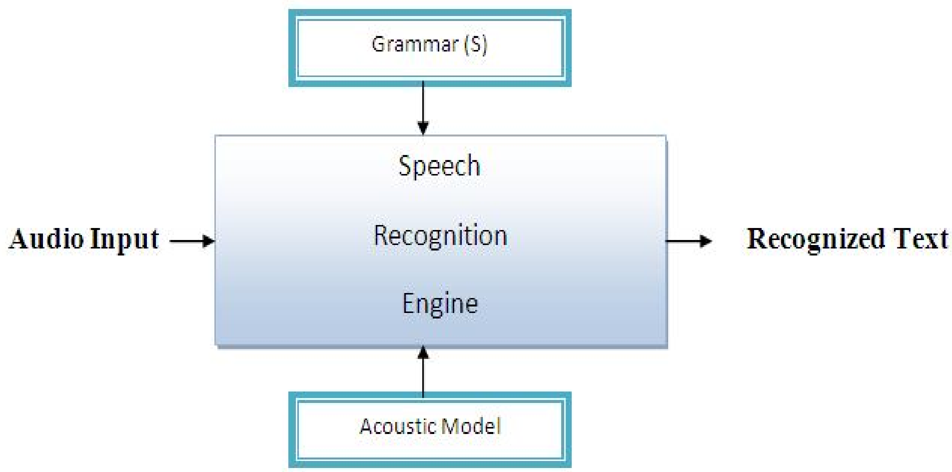


Figure 1: Speech Recognition Engine Components

1- Language Model or Grammar: A Language Model is a file containing the probabilities of sequences of words. A Grammar is a much smaller file containing sets of predefined combinations of words. Language Models are used for 'Dictation' applications, whereas Grammars are used as desktop 'Command and Control' applications.

2- Acoustic Model: Contains a statistical representation of the distinct sounds that make up each word in the Language Model or Grammar. Each distinct sound corresponds to a phoneme.

Speech Recognition Engine also uses a Software program that is called Decoder, which takes the sounds spoken by a user and searches the Acoustic Model for the equivalent sounds. When a match is made, the Decoder determines the phoneme corresponding to the sound. It keeps track of the matching phonemes until it reaches a pause in the users' speech. It then searches the Language Model or Grammar file for the equivalent series of phonemes. If a match is made it returns the text of the corresponding word or phrase to the calling program.

2. ALGORITHMS AND MODELS

2.1. Dynamic Time Warping:

The Dynamic Time Warping (DTW) is an algorithm, it was introduced in 1960s [10]. It is an important and aged algorithm was used in speech recognition systems known as Dynamic Time Warping algorithm [7] [12] [14], It is used to measure the resemblance of objects/Sequences in the form of speed or time. For example similarity would be detected in running pattern where in film one person was running slowly and other person was running fast. This algorithm can be applied to any data; even data is graphics, video or audio. It analyzes data by turning into a linear representation.

This algorithm is used in many areas: Computer animation, Computer vision, data mining [13], online signature matching, signal processing [9], gesture recognition and speech recognition [2].

2.2. Hidden Morkov Model

It is modern general purpose algorithm. It is widely used in speech recognition systems because of that statistical models are used by this algorithm, which creates output in the form of series of quantities or symbols. It is based on statistical models that output a series of symbols or quantities [3].

2.3. Neural Networks

Neural networks were created in the late 1980s. These were emerging and an attractive acoustic modeling approaches used in Automatic Speech Recognition (ASR). From the time then these algorithms have been used in various speech based systems such as phoneme categorization [1] speaker adaptation and isolated word recognition [8]. These algorithms are attractive recognition models for speech recognition because they formulate no assumptions as compared to Hidden Markov Models regarding feature statistical properties. This algorithm is used as preprocessing i.e; dimensionality reduction [6] and feature transformation for Hidden Markov Model based recognition.

3. PROPOSED WORK

The research is concentrated on the four language models/grammars, which are implemented in Text Editor through Voice. Those models/grammars are;

1) Dictionary
2) HTML (Hypertext Markup Language)
3) IDE (Integrated Development Environment)
4) Special Characters (S.Characters)

From these four language models/grammars, one is used as ‘Command & Control’ purpose and three are used for ‘Dictation’ purpose. Their classification is given in Figure# 2.

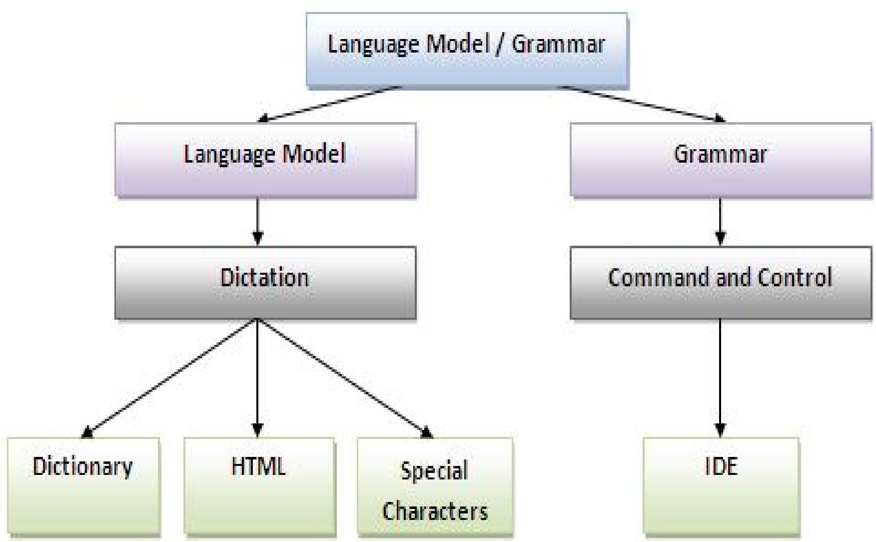


Figure 2: Classifications of Language Model / Grammar

4. IMPLEMENTATION OF PROGRAMMED LANGUAGE MODELS & GRAMMAR

As discussed in introduction section that Speech Recognition Engine requires two types of files to recognize inputs. First is the Language/Grammar model and the second is acoustic model. So we have created three language models and one grammar In the Figure# 3 we have shown the implementation of language models and grammar model in speech recognition engine.

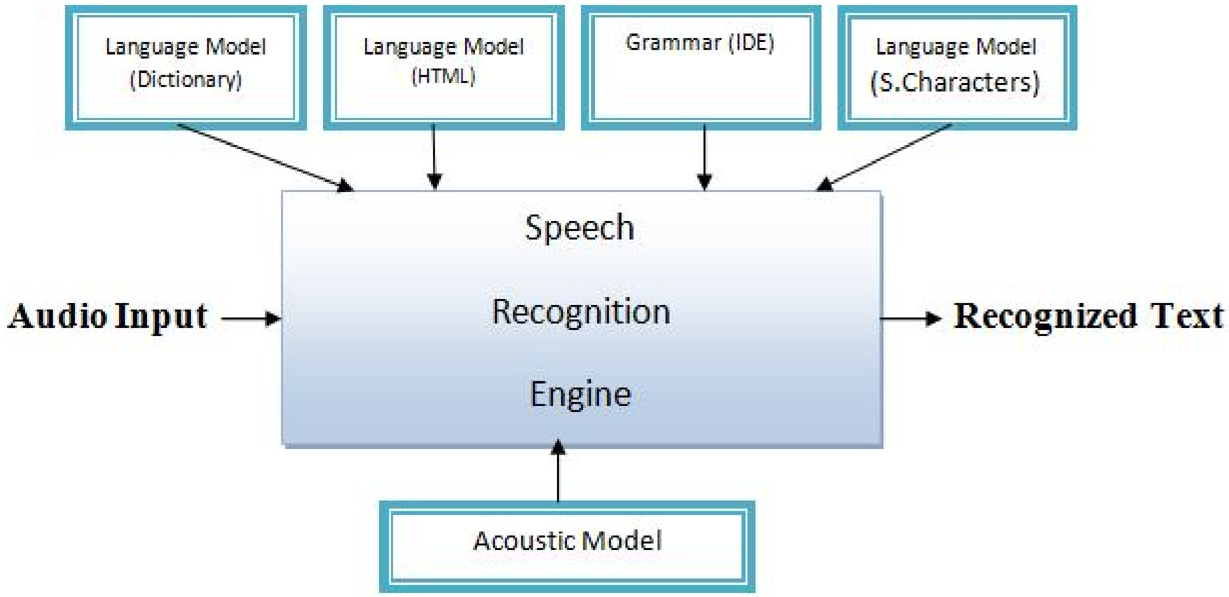


Figure 3: Implementation of Language Models & Grammar

5. APPLICATION SNAPSHOTS AND RESULTS

Figure #4 is GUI (Graphical User Interface) of our designed application. In the left side of application we have given four MIC icons means these are functions in order to use and analyze language models and grammar.

.

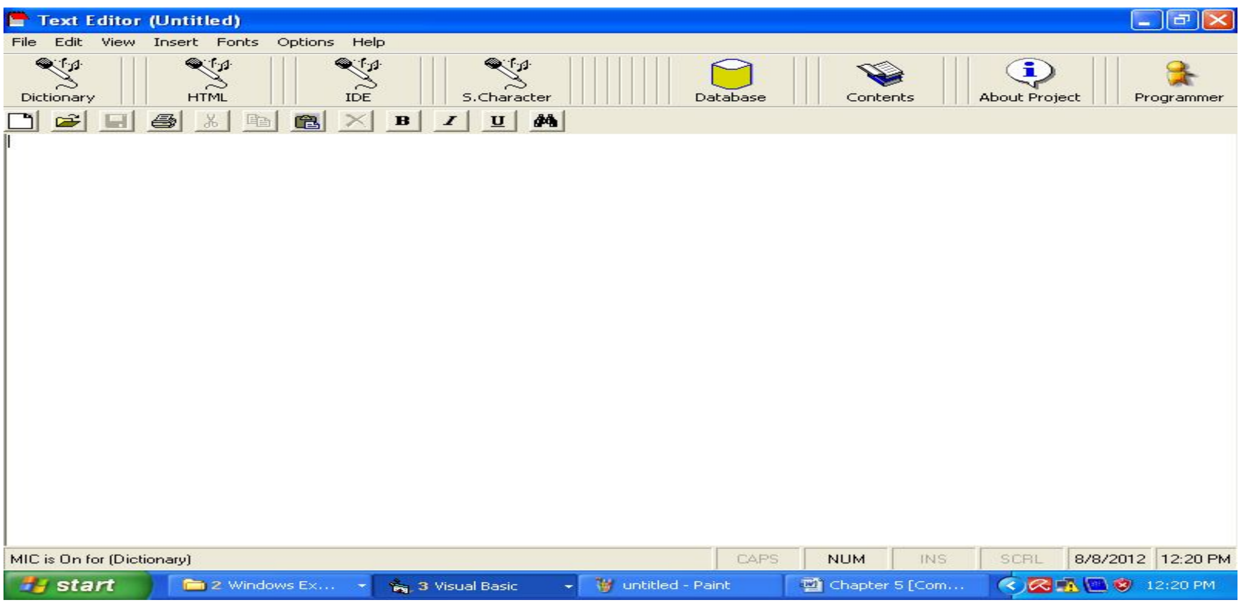


Figure 4: (Text Editor Through Voice) Editor Window

5.1. Dictionary: This is Language Model used for dictation purpose in which a user can add and use words, phrases, names and sentences in documents. 10000 words are already stored in dictionary database. Figure #5 shows recognizing of some words and letters in document area which are inserted by speaking into MIC.

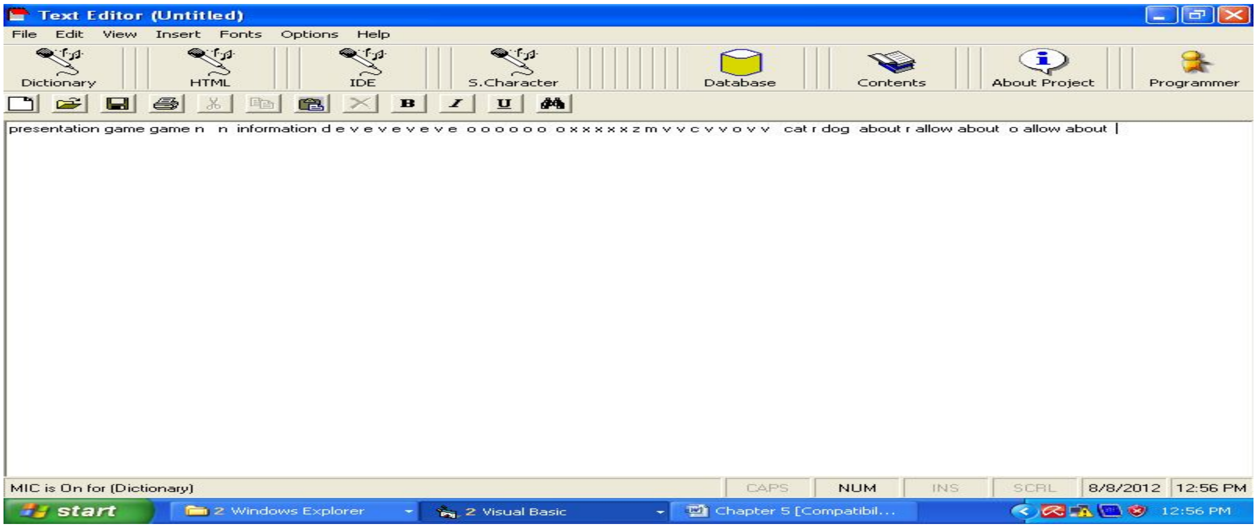


Figure 5: (Editor Window) Editing by MIC (Selected function is Dictionary)

5.2. HTML: This is Language Model used to create simple web scripts based on dictation. Words and phrases for their corresponding HTML Tags are given in table no: 1. and figure #6 shows simple web script created by speaking their phrases into MIC.

Phrases	Opening Tags	Phrases	Closing Tags
HTML	<HTML>	Close HTML	</HTML>
HEAD	<HEAD>	Close HEAD	</HEAD>
TITLE	<TITLE>	Close TITLE	</TITLE>
Body	<Body>	Close Body	</Body>
Image	<Image>	---	---
B	<B>	Close B	</B>
I	<I>	Close I	</I>
U	<U>	Close U	</U>
Center	<Center>	Close Center	</Center>
Font	<Font>	Close Font	</Font>
HR	<HR>	Close HR	</HR>
BR	 	Close BR	</BR>
P	<P>	Close P	</P>
Table	<Table>	Close Table	</Table>
TH	<TH>	Close TH	</TH>
TR	<TR>	Close TR	</TR>
TD	<TD>	Close TD	</TD>
H1	<H1>	Close H1	</H1>
H2	<H2>	Close H2	</H2>
H3	<H3>	Close H3	</H3>
H4	<H4>	Close H4	</H4>
H5	<H5>	Close H5	</H5>
H6	<H6>	Close H6	</H6>
Sub	_{	Close Sub	}
Sup	^{	Close Sup	}
Marquee	<Marquee>	Close Marquee	</Marquee>
Frame	<Frame>	Close Frame	</Frame>
Frameset	<Frameset>	Close Frameset	</Frameset>
Form	<Form>	Close Form	</Form>
Input	<Input>	---	---
Select	<Select>	Close Select	</Select>
Option	<Option>	---	---
Text Area	<Textarea>	Close Text Area	</Textarea>

Table No 1: List of Phrases and HTML Tags

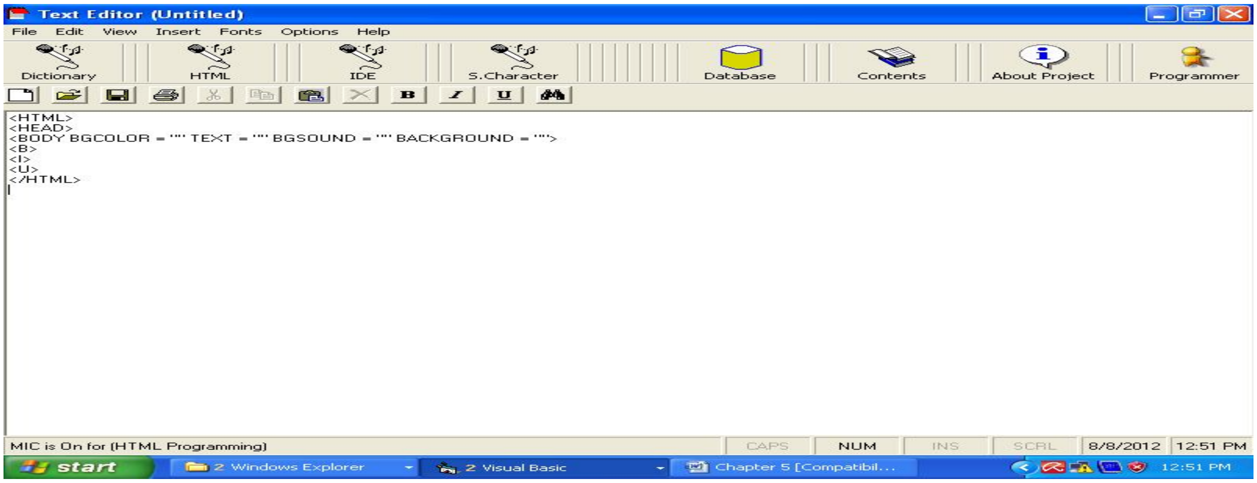


Figure 6: (Editor Window) Editing by MIC (Selected function is HTML)

5.3. IDE: This is Grammar used as command and control purpose. Phrases and descriptions are given in table no: 2 figure #7 shows date and time function is called by speaking corresponding phrase into MIC.

List of Phrases	**Description of Phrase**
New	To Open new document
Open	To Open saved document
Save	To Save Document
Save As	To Save document with new name
Print	To Print document
Exit	To Exit Text Editor
Delete	To Delete selected text
Cut	To Cut selected text
Copy	To Copy selected text
Paste	To place cut or copied text
Find	To Search text from document
Replace	To Replace document
Select All	To Select All Text
Time	To Insert time in document
Tool Bar	To Call tool bar function
Status Bar	To Call status bar function
Standard Buttons	To Call standard buttons function
Date and Time	To Insert date and time in document
Bold	To change the format of text as Bold
Italic	To change the format of text as Italic
Underline	To change the format of text as underline
Font	To Call font function
Color	To Call color function
Dictionary	To call Dictionary function
HTML	To call HTML function
IDE	To call IDE function
Special Characters	To call special character function
Database	To call database wizard function
De Activate	To Off MIC
Capital Characters	To call capital character

	function
Small Characters	To call small character function
About Me	To know about Application Developer
About Project	To know about Project Description
Contents	Help and Index

Table No 2: List of phrases to control IDE

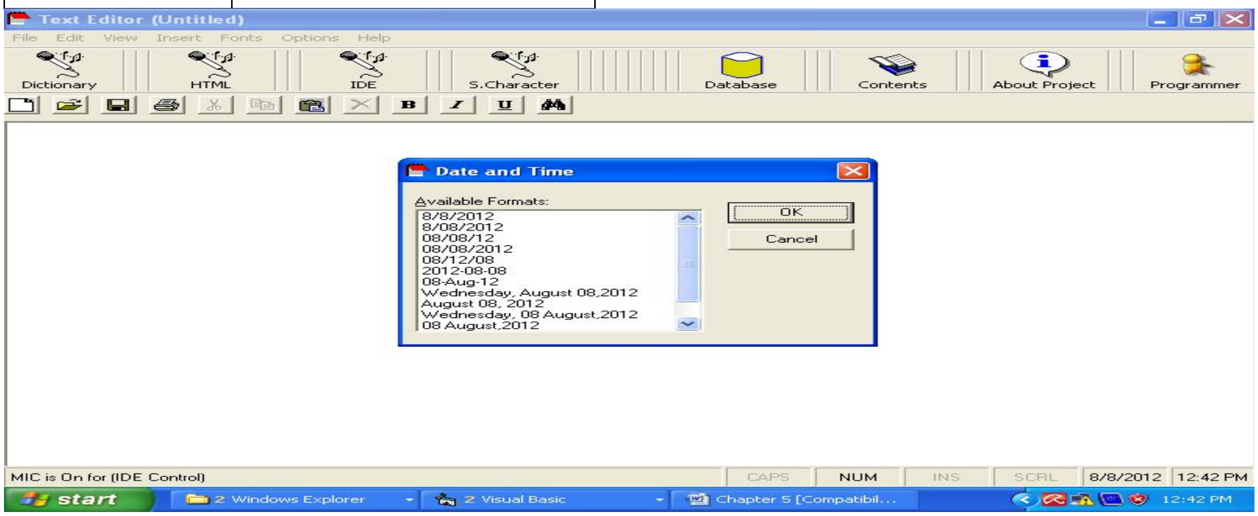


Figure 7: (Editor Window) Editing by MIC (Selected function is IDE)

5.4. Special Characters: This is Language Model used for dictation purpose. Users can insert special characters and numbers in documents. Phrases and descriptions are given in table no: 3 and figure #8 shows some special characters and numbers in document area which are inserted by speaking into MIC.

List of Phrases	Description
Less than	To insert (<) sign in document
Greater than	To insert (>) sign in document
Dot	To insert (.) sign in document
Comma	To insert (,) sign in document
Colon	To insert (;) sign in document
Semi colon	To insert (:) sign in document
Single quote	To insert (‘) sign in document
Double quote	To insert (“) sign in
	document
Question mark	To insert (?) sign in document
Steric	To insert (*) sign in document
And	To insert (&) sign in document
Percent	To insert (%) sign in document
Slash	To insert (/) sign in document
Back slash	To insert (\) sign in document
Hash	To insert (#) sign in document
Dollar	To insert ($) sign in

	document
Dash	To insert (-) sign in document
Underscore	To insert (_) sign in document
Exclamation	To insert (!) sign in document
Addition	To insert (+) sign in document
Subtraction	To insert (-) sign in document
Zero	To Insert (0) sign in document
One	To insert (1) sign in document
Two	To insert (2) sign in document
Three	To insert (3) sign in document
Four	To insert (4) sign in document
Five	To insert (5) sign in document
Six	To insert (6) sign in document
Seven	To insert (7) sign in document
Eight	To insert (8) sign in document
Nine	To insert (9) sign in document
Back	To call the function of (back space) key
Insert	To call the function of (insert) key
Delete	To call the function of (delete) key
Home	To call the function of (home) key
End	To call the function of (end) key
Page up	To call the function of (page up) key
Page down	To call the function of (page down) key

Table No 3: List of phrases for special characters and functions

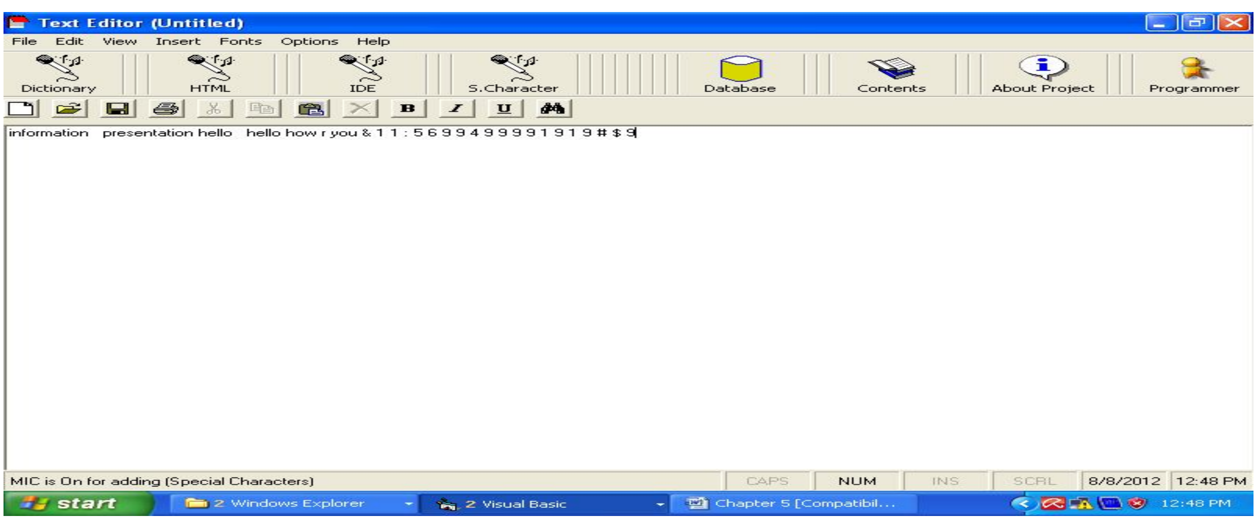


Figure 8: (Editor Window) Editing by MIC (Selected function is S.Characters)

6. CONCLUSION

It is learnt through this project that the professional work is necessary to be carried out in the software industry. It is proved through study to develop Speech based application named Text Editor Through Voice. Here are designed two types of Language Models/Grammars, and have classified them as dictation and command & control grammars. Further a concept have portrayed that computer programmers can create simple web scripts through the speech base process. It is concluded through Graphical User Interface (GUI) and outputs, to make it possible, to create web scripts via speaking commands. The study is implemented in designed grammar in speech recognition engine in order to prove the solution, which is technically feasible.

The work on this application is oriented to direction as a commercial system. Hidden Markov Model is used in usually speech recognition software, which integrates an acoustic model, a large vocabulary file. One of the software version used for the test phase is equipped with a vocabulary gathering more than 10000 current and specialized words. Ideally, the use of a voice recognition system, really speaker independent, like the Microsoft word processor, we want to improve the robustness of the designed language model and grammar in the whole application. Best results will be selected among different speech recognition engines after implementation according to a framework working at the same time. The software industry is using the machines but it is little tried to create some applications for commercial purposes. Our future work is oriented in this direction as commercial systems. In this scenario it is tried to level best to develop Speech Based Text Editor, which is working properly and needed more improvements as a successful commercial product. For example:

- This application cannot understand any input if was spoken in other than English language. There is acute need to develop an Editor for Sindhi and Urdu languages.
- This application needs perfect pronunciation, sound proof of environment and having no noise.
- There is need to develop the same application in .Net Framework for latest equipments and easy to access on each platform.

ACKNOWLEDGEMENT

The instigator is grateful to the University of Sindh, Jamshoro, for providing its economical help for the research project

REFERENCES

[1] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, (1989) "Phoneme recognition using time-delay neural networks," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 37, pp. 328-339.

[2] C. Myers, L. Rabiner, and A. Rosenberg, \Performance tradeo_s in dynamic time warping algorithms for isolated word recognition," Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on, vol. 28, no. 6, pp. 623{635,1980.

[3] Goel, V.; Byrne, W. J. (2000). "Minimum Bayes-risk automatic speech recognition". *Computer Speech & Language* 14 (2): 115–135. doi:10.1006/csla.2000.0138. Retrieved 2011-03-28.

[4] H. Dudley, *The Vocoder*, Bell Labs Record, Vol. 17, pp. 122-126, 1939.

[5] H. Dudley, R. R. Riesz, and S. A. Watkins, *A Synthetic Speaker*, J. Franklin Institute, Vol. 227, pp. 739-764, 1939.

[6] Hongbing Hu, Stephen A. Zahorian, (2010) "Dimensionality Reduction Methods for HMM Phonetic Recognition," ICASSP 2010, Dallas, TX

[7] Itakura, F. (1975). Minimum Prediction Residual Principle Applied to Speech Recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 23(1):67-72, February 1975. Reprinted in Waibel and Lee (1990).

[8] J. Wu and C. Chan,(1993) "Isolated Word Recognition by Neural Network Models with Cross-Correlation Coefficients for Speech Dynamics," IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, pp. 1174-1185.

[9] M. Muller, H. Mattes, and F. Kurth, \An e_cient multiscale approach to audio synchronization," pp. 192{197, 2006.

[10] R. Bellman and R. Kalaba, \On adaptive control processes,"Automatic Control, IRE Transactions on, vol. 4, no. 2, pp. 1{9,1959.

[11] S. A. Zahorian, A. M. Zimmer, and F. Meng, (2002) "Vowel Classification for Computer based Visual Feedback for Speech Training for the Hearing Impaired," in ICSLP 2002.

[12] Sakoe, H. and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for

Spoken Word Recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 26(1):43-49, February 1978. Reprinted in Waibel and Lee (1990).

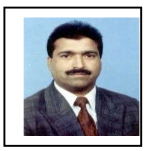
[13] V. Niennattrakul and C. A. Ratanamahatana, \On clustering multimedia time series data using k-means and dynamic time warping," in Multimedia and Ubiquitous Engineering, 2007. MUE '07. International Conference on, 2007, pp. 733{738.

[14] Vintsyuk, T. (1971). Element-Wise Recognition of Continuous Speech Composed of Words from a Specified Dictionary. *Kibernetika* 7:133-143, March-April 1971.

BIOGRAPHY

M.phil(Computer Science), MCS (Computer Science), BCS (Computer Science), CCNA (Cisco Certified Network Associate) Certified, Working as Teaching Assistant, Research Interest in Speech Recognition, Computer Network Security, Database Designing, Algorithm Designing and Computer Programming.

Ph.D (Environmental Sciences), M.Phil (Environment Sciences), BS (Information Technology) Research Interest in environmental atmospheres, in particular interpreting of the Earth's atmosphere and surface by satellites; will offer most effective ways to monitor and study global change. I am also interested in the application of Climate and Environmental modeling for the current climate and environmental fashion and analytical humans' possible impacts on climate.



Ph.D (Computer Science & Technology), Research Interest in Distributed middleware, Distributed Mobile Computing, Formal language models and implementation techniques, Web-Based Interface, Human Computer Interaction (HCI), Enterprise Services Integration and Organizations Collaboration of Distributed Systems and Net6works.

Ph.D (Computer Science & Technology), Research Interest in Advanced Software Engineering. Collaborative Software Development Methods, Agile Software Development, Distributed Software Development, Open Source Software Development, Commercial-off-the Shelf (COTS) based Software Development, Machine Learning.

Ph.D. (Computer Science), Research Interest in Service-oriented Computing, Semantic Web Technologies, Knowledge Management, Software Engineering.