

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

English to Paite Translator using Speech Recognition Technology

Chingmuankim

M.Tech Scholar, Dept. of CSE, NERIST, Itanagar, India

ABSTRACT: English to Paite Translator is an application that converts the input Speech in English to Paite Language in the form of text. When the client utters a sentence in English, the Speech recognizer searches the matching input speech in the database and the corresponding sentence in Paite is retrieve by the SQL Server and is displayed on the screen. Unmatched Speech results in displaying an error message. This System is based on Speaker Independent Speech Recognition System where the system is independent on the voice of the client but is dependent on the uttered words. The application is implemented using C-sharp on Visual Studio and the data are stored in SQL database.50 sentences in English are tested. The system while testing is found to perform very well. It is observed after testing that the time taken by the system to respond to the corresponding input Speech is approximately 0.027hrs.

KEYWORDS:Speech Recognition, Integrated Development Environment, Hidden Markov Model, Terminal

I. INTRODUCTION

English to Paite Translator allows you to utter speech in English and later convert the input speech into Paite.Language being one factor that creates bonding between humans, it is therefore required to learn different languages. Hence this system is developed for people who are interested in learning Paite language. When the Client utters in English the Paite sentence he/she wants to know, the Speech Recognition Engine after recognizing the input speech returns its corresponding meaning to the Screen in machine-readable form.Paite is a Tibeto-Burman Language spoken in some parts of Manipur province of Northeast India and Myanmar. The system is Independent on who the Speaker is. It is dependent on the pronunciation of words.

Back in old days, we have to tolerate the burden of typing and is also time consuming, but with the introduction of Speech Recognition Technology which is also known as a Hands-free Technology, we are now free from the burden of typing which not only makes our life easier but also saves time. Speech Recognition System can either operate directly on the system or on the server. A terminal that has a Speech Recognition Engine on the system operates directly on the terminal while the terminal that uses a server transmits the vibrations of the Speech Signals to the server which runs on a Speech Recognition Engine and returns the corresponding text result in Paite to the terminal.

The work is implemented on Microsoft Visual Studio using the language C-Sharp. It is an Integrated Development Environment (IDE) from Microsoft. It has an inbuilt Speech Recognition engine which provides the Speech Recognizer to recognize the user's speech and integrate it with the inbuilt simulator.It is used to develop computer programs, as well as desktop applications, websites, web apps, web services and mobile apps. C-Sharp is purely an Object Oriented Programming Language.It is a general-purpose language designed for developing apps on the Microsoft platform and requires the .NET framework on Windows to work. C-Sharp is oftenthought of as a hybrid that takes the best of C and C++ to create a truly modernized language. Although the .NET framework supports several other coding languages, C-Sharp has quickly become one of the most popular.

C-Sharp can be used to create almost anything but is particularly strong at building Windows desktop applications and games. C-Sharp can also be used to develop web applications and has become increasingly popular for mobile developmenttoo. Cross-platform tools such as Xamarin allow apps written in C-Sharp to be used on almost any mobile device.This work is based on Desktop application using the language C-Sharp.

The paper is further organized into the following Sections. Section II describes the related work giving details about the Literature review. Section III gives an overview of the system describing the proposed methodology, flow of work,

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

flowchart, user interface, Database creation and structure of the table. The flow diagram represents the step of the algorithm. Section IV presents experimental results showing the results of images tested. Finally, Section V presents the conclusion.

II. RELATED WORK

[1] developed an android application to convert Speech to Speech. The application converts English Speech to Tamil. The process of Speech to Speech translation involves three core technologies. These are Speech Recognition Technology, Machine Translation and Speech Synthesis Technology. In Speech Recognition Technology, the system recognizes the input English Speech via Speech Recognition device and converts it into text. The Machine Translation Technology translates the text in a certain language into text in another language and Speech Synthesis technology converts the translated text into a Speech. 60 words are tested and voice play upto 20 words using an embedded system.

[2]implemented Speech Recognition in Malay language, 90 Malay phrases were tested by 10 speakers from both genders in different contexts. The result shows that the overall accuracy is satisfactory as it is 92.69%. The result was calculated using Confusion Matrix where the Sensitivity and Specificity curve was plotted and is found to be reliable.

[3]implemented a large vocabulary based on a large dataset in order to improve Speech Recognition performance and demonstrates the effectiveness of this approach. Language processing technique is applied to the result of Speech Recognition. In this work, they have developed a sample application to support user operations. The Article also describes the Speech Recognition Technology and development of the application.

[4] aims at developing a speech supportive system for Dysarthria speakers. Dysarthria is a speech disorder caused by muscle weakness which makes it hard for a person to talk People have a hard time understanding what they say as muscles are used to talk. These muscles include those in our face, lips, tongue and throat. Its objectives are to identify the articulatory errors of each Dysarthria Speakers. In this work, they have developed a Speech Recognition System that corrects the errors in dysarthria speech by incorporating the findings from the first-fold using a speaker specific dictionary. They have developed an HMM-based Speaker adaptive Speech Synthesis system that synthesizes the error corrected text for each dysarthria speakers retaining their identity. It was tested for 10 Dysarthria Speakers. Articulatory errors are analyzed and identified. Errors are rectified and are synthesized into Speech. This synthesized speech is checked for its intelligibility and naturalness using Mean Opinion Score (MOS). For adaptation, out of 74 utterances of each dysarthria speakers which is fair.

[5] presented an AISHELL-1 Corpus. AISHELL-1 is an open source Mandarin ASR corpus suitable for conducting Speech Recognition Research and building Speech Recognition Systems. A character error rate of 12.64 %, 12.42% and 10.90% is obtained.

This paper, however is focussed more on how Speech Recognition Technology can be used in Visual Studio to translate the input English Speech into Paite Language using C-Sharp, rather than focussing on Machine Translation and Speech Synthesis. It provides a more detailed description of the programming language used and a little background of Paite Language and the flow of work.

III.OVERVIEW OF THE SYSTEM

Speech Recognition requires two basic components in order to recognize the Speech. These are the Acoustic model and Language model. The Acoustic Model contains the statistical representation of phonemes or sequence of words. These statistical representations are called Hidden Markov Model (HMM). Phonemes are distinct sounds that make up a word. They (Phonemes) can have the same spelling but different pronunciation and meaning. This Model represents the language acoustic sounds which recognizes the characteristics of a certain user's speech patterns. Each phoneme has its own HMM. HMM make use of state machines. In HMM, the states are hidden and can be predicted from the evidence. The sequence of words are the hidden states. A hidden Markov model for a sequence of words is made by concatenating the individual trained Hidden Markov Models for the separate words. The figure below demonstrates an HMM for the word "need".

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

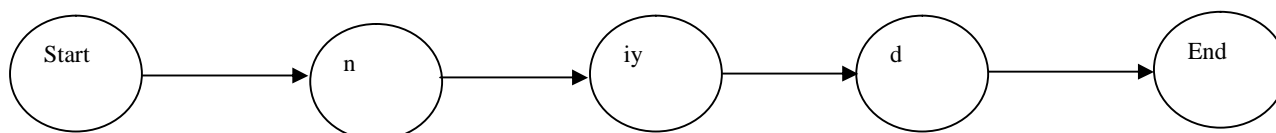


Fig: 1 Word model for Need

The language model is a file used by Speech Recognition Engine to recognize Speech. Language Model contains a large list of words. It provides context to distinguish between words and phrases that sound similar. For eg: The word "industries" and the phrase "in the streets" sounds similar but totally different meanings.

The Quality of a speech recognition system depends on the quality of the acoustic models and language of recognizer and the effectiveness of the algorithm for processing input sounds and searching across its models. Grammar includes the list of words and phrases to be recognized. Speech Recognition grammar is a set of rules or constraints that define what a speech recognition engine can recognize as meaningful input. It is the most important part of the Speech Recognition System as it limits the number of recognized words and phrases. If the system is set up with a grammar file to recognize the word "I need food" whose phonemes are "ay n iy d f oo d", the Speech Recognition Engine follow the following steps. Initially, the Speech decoder listens to the distinct sounds uttered by the client and then searches for the matching HMM in the Acoustic model. When it finds a matching HMM in the Acoustic model, the decoder takes note of the phoneme. The decoder keeps a record of the matching phoneme until there is a pause on the Speech. When reaching pause, the decoder looks up the matching series of phonemes it heard and then display the corresponding output. In this work, the words and phrases are loaded from the database to the grammar by the Speech Recognition engine.

In order to create a Speech Recognition System, "System. Speech" need to be first referenced. "System. Speech" includes special recognition modes that allow you to provide text instead of audio to the speech recognition engine. The System.Speech.Recognition namespace has the functionality to monitor the speech input, create the speech recognition grammars that produce both literal and semantic recognition results and manage speech engines. With this input functionality, the application can monitor the state, level and format of the input signal and receive notification about problems that might interfere with successful recognition. The Grammar Builder Class provides a mechanism for programmatically building the constraints for a speech recognition grammar. SpeechRecognitionEngine and SpeechRecognizer classes, can select the Speech Recognition Engine and configure the input, perform start and stop recognition, load and unload grammars and modify the properties of the speech recognition engine that affect recognition. The event handler for the Speech recognized events is invoked when the recognizer receives input that matches one of its Speech Recognition grammar. The Speech Recognised event is used to display the successfully recognized phrases. Spoken input such as "Good morning" will trigger the SpeechRecognised event because it is predefined in the grammar.

The User interface has six buttons. These are Speak, Stop, and Clear, Quit, Start and another Stop button for the Stopwatch. The Speak button enables the machine to recognize the Speech while the Stop button disables the Speech recognition engine to recognize Speech. The textbox gets cleared when the clear button is pressed. Quit button exits the windows application. Start button starts the timer and the end button stops the timer. The Stopwatch has labels to display time in Hours, Minutes, Seconds and Centiseconds. Flowchart of the program is given below in Fig 2 and 3.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

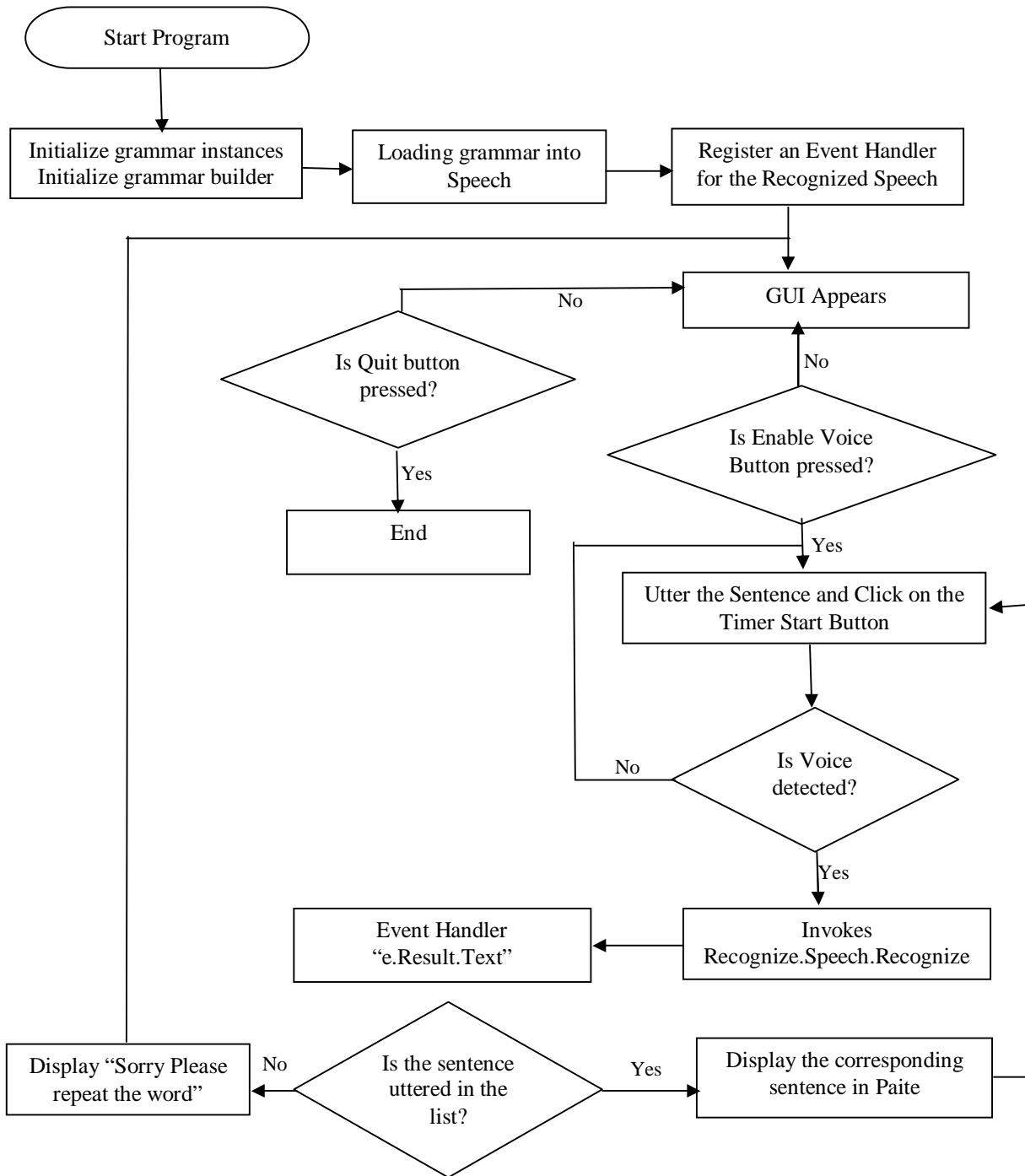


Fig. 2. Program Flowchart

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

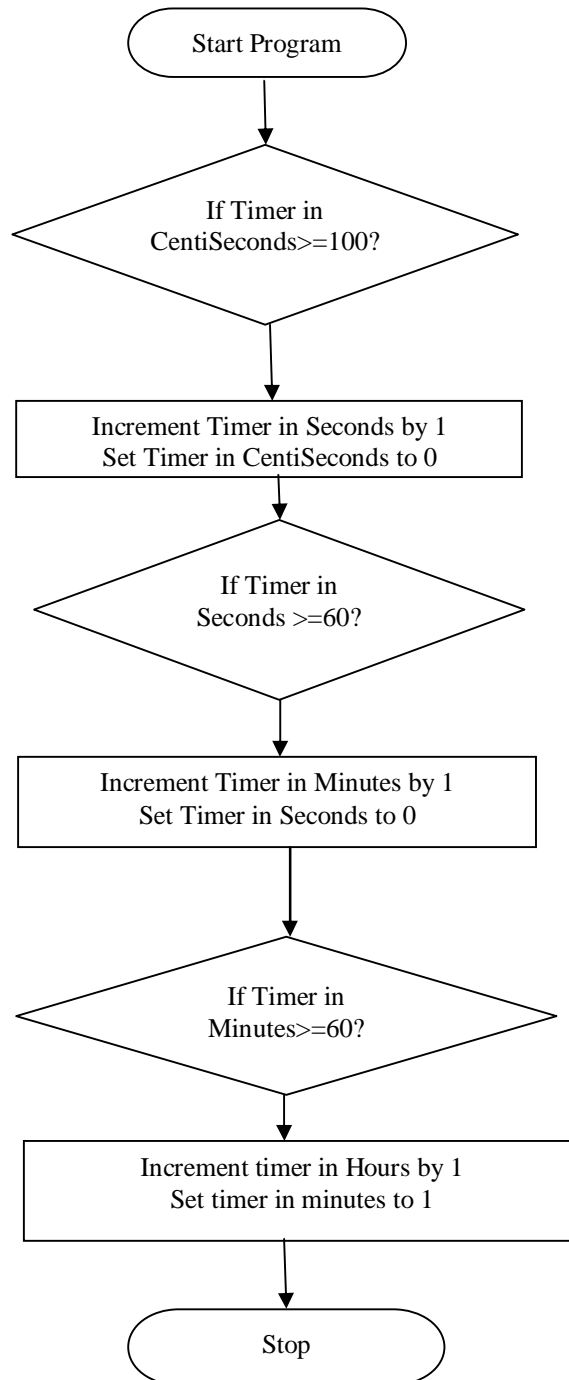


Fig. 3. Flowchart of Stopwatch

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

The System has an extra feature like Stopwatch (Timer) to count the time interval of the System testing time. Fig 3 shows the flowchart of the Stopwatch. The start button enables the system to start counting the time while the Stop button disabled the time. The timer has a label to display the time in Hour, Minute, Seconds and Centiseconds. When the time in Centiseconds is greater than or equal to 100, the time in seconds is incremented by one value each time and when these seconds cross or equals 60, the minute is incremented based on the seconds' timer. Finally, the hour gets updated and gets incremented when the minute timer is greater than or equal to 60.

Name	Owner	Type	Created_datetime	
1	tb1	dbo	user table	2019-02-05 08:23:24.957

Column_name	Type	Computed	Length	Prec	Scale	Nullable	Trim TrailingBlanks	FixedLenNullInSource	Collation
1	English	nvarchar	no	900		no	(n/a)	(n/a)	SQL_Latin1_General_CP1_CI_AS
2	Paite	nvarchar	no	900		no	(n/a)	(n/a)	SQL_Latin1_General_CP1_CI_AS

Identity	Seed	Increment	Not For Replication	
1	No identity column defined.	NULL	NULL	NULL

RowGuidCol	
1	No rowguidcol column defined.

Data located on filegroup	
1	PRIMARY

index_name	index_description	index_keys	
1	PK_tb1_6E814911FE8354D2	clustered, unique, primary key located on PRIMARY	English

constraint_type	constraint_name	delete_action	update_action	status_enabled	status_for_replication	constraint_keys	
1	PRIMARY KEY (clustered)	PK_tb1_6E814911FE8354D2	(n/a)	(n/a)	(n/a)	(n/a)	English

Fig. 4. Table Structure

Fig 4 includes the Decrypt_Default_name column indicating the decrypt default name for the column. As shown on the figure, name of the table is "tb1" and is created on "05-02-2019". There are two columns in the table. These are "English" and "Paite". Both are having nvarchar datatypes. The "N" in NVARCHAR means Unicode. nvarchar is nothing more than a varchar that supports two byte characters. This datatype can store a mixture of both English and non-English Symbols.

	name	rows	reserved	data	index_size	unused
1	tb1	50	16 KB	8 KB	8 KB	0 KB

Fig. 5. Database Size

Fig 5 shows the amount of available database space. The table "tb1" holds 50 total data. The figure also gives an overview of the reserved space of the table, size of the data, index size and the unused space. Reserved space is the amount of space that has been allocated to all the tables and indexes created in the database. The data space gives the amount of space occupied by the data. The size of an index depends on four things. These are the number of entries or rows, number of columns in the key, size of the column values and the number of similar key values. The Unused space is the space on the disk that does not have any data on it or a reserved space that is not yet used by the existing tables and indexes. The sum of the values in the unused, index size and data columns should equal the figure in the reserved column.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

IV. EXPERIMENTAL RESULTS

The system is tested with 50 sentences in English which after being recognized by the recognition engine displays the text on the screen and its corresponding output in Paite Language is displayed on the screen as shown in Fig 6. It is found that the time taken by the system to translate fifty English Speech to Paite is one minute, thirty-seven seconds, seventy-five centiseconds which is approximately 0.027 hours. The System is tested without using a microphone. It can be observed from the time taken by the system that it is very efficient.

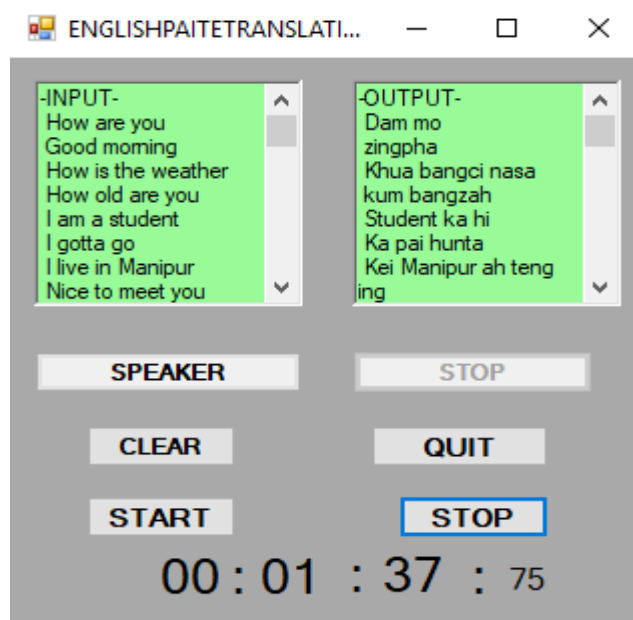


Fig. 6. Tested data

V. CONCLUSION

In this work, Speech to text translation from English to Paite on Visual Studio is implemented. The adopted algorithm successfully recognize the input speech which also converts the input speech to text in a different language (Paite). 50 English speech is successfully tested and the result is found to be good. C-Sharp programming language performs fast and accurate. The System shows a good performance within a short period of time. The result is reliable and accurate. An extension of the work in the future would be to apply Machine Learning Techniques by using Machine Translation to automatically translate the language into different languages making use of the Cloud translation API.

REFERENCES

- [1] V.J.Poornakala,A.Maheshwari,"Automatic Speech-Speech Translation System from English to Tamil Language",IJIRSET,vol. 5, Issue 3, March2016.
- [2] Rami Ali Al-Khulaidi and Rini Akmeliawati, "Speech to Text translation for Malay Language", IOP Conference, 2017.
- [3] Shinya Iizuka, Kosuke Tsujino, Shin Oguri, Hirotaka Furukawa, "Speech Recognition Technology and Applications for improving Terminal Functionality and Service Usability", NTT DOCOMO Technical Journal, vol. 13, No. 4, 2012.
- [4] M.Dhanalakshmi, T.A.Mariya Celin, T. Nagarajan1, P.Vijayalakshmi1, "Speech-Input Speech-Output Communication for Dysarthria Speakers using HMM-Based Speech Recognition and Adaptive Synthesis System", Springer Journal, vol 37, No 2, 2018
- [5] Hui Bu, Beijing Shell Shell Technology Co. Ltd, Independent Researcher*, Beijing, China "An open-source mandarin Speech Corpus and a Speech Recognition Baseline," IEEE 20th Conference of O-COCOSDA, 2017
- [6] T.T.S.a.S.H.S.Salleh, "Corpus-based Malay Text-to-Speech Synthesis System" 14th Asia-Pacific Conference on Communications, Tokyo,2007

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 1, January 2019

- [7] B. Ramani, Lilly Christina, G Anushiya Rachel, V Sherlin Solomi, Mahesh Kumar Nandwana, Anushka Prakash, Aswin Shanmugam S, Raghava Krishnan, SP Kishore. K Samudravijaya, P Vijayalakshmi, T Nagarajan and Hema A Murthy, "A common attribute based unified HTS framework for Speech Synthesis in Indian Languages", 8th ISCA Speech Synthesis Workshop, Barcelona, Spain, 2013.
- [8] M.TELMEM, Y.GHANOU "Estimation of the Optimal HMM Parameters for Amazigh Speech Recognition System using CMU-Sphinx" The First International Conference on Intelligent Computing in Data Sciences, 13 March 2018.
- [9] M. Dhanalakshmi, T.A. Mariya Celin, T.Nagarajan, P.Vijayalakshmi, T.Nagarajan, P. Vijayalakshmi, "Speech-Input Speech-Output Communication for Dysarthria Speakers Using HMM-Based Speech Recognition and Adaptive Synthesis System", Journal, Circuit Systems, and Signal Processing, Volume 37 Issue 2 khauser Boston Inc. Cambridge, MA, USA, 1 February 2018.
- [10] Hinton, G. et al. "Deep neural networks for acoustic modeling in speech recognition", IEEE Signal Processing Magazine 29, 82–97 (2012).
- [11] Sangmi Shin, Eric T. Matson, Jinok Park, Bowon Yang, Juhee Lee, Jin-Woo Jung, "Speech-to-Speech Translation Humanoid Robot in Doctor's Office", 6th International Conference on Automation, Robotics and Applications, 2015.
- [12] Begum A, Askari S.M.S., Sharma U, In Panigrahi C., Pujari A., Misra S., Pati B., Li KC. (eds) "Text-to-Speech Synthesis System for Mymensinghiya Dialect of Bangla Language, Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing", vol 714. Springer, Singapore, 2019