International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 3, Issue 1 (January 2014), PP. 39-41

A Novel Model for Speech to Text Conversion

Deepa V.Jose, Alfateh Mustafa, Sharan R
Department of Computer Science, Christ University, Hosur Road, Bangalore-23

Abstract:- Gaining fluency in any language is a cumbersome task. To make it easy so many software's are existing now a days. Our aim is to develop a software that enhances the user's way of speech through correctness of pronunciation following the English phonetics. This software allows one to learn, judge and recognize their potential in English language. It also facilitates an extra add-on feature which nourishes the user's communication skills by an option of text to speech conversion conversion also. Enhancing the existing algorithms of speech to text to improve the quality of the output is under consideration.

Keywords:- cryptography, natural language processing, phonetics, text to speech conversion, speech to text conversation,

I. INTRODUCTION

Learning can never have an end, especially when it is related to learn a language. There is always an addition at every interval and it is very vast, which makes is not very easy for anyone to master a language easily. Text to speech [TTS] technology is the process wherein the computer is made to speak. It uses the concepts of natural language processing. The Speech synthesizer converts the audio input into the text form and processes the text for further learning modules.

Traditionally, acquiring the knowledge to master a language was done through various language-learning related textbooks. Moving along this approach of learning was very difficult because of the cost factor. The existing system deals with various dictionaries, which implements dictation of words with correct pronunciation. It supports the operation, only for a set of words, which are available in the dictionary. The availability of such software doesn't eradicate the problem, which is being discussed in the picture. The system speaks out the selected word, which the user wishes to listen to. It also includes various features, which suggests the adjective and verb form of the word, which can indirectly increase the knowledge of English Language. This method implied no fun and it was very monotonous as well.

The current system is focused more on polishing the pronunciation from better to best and not focused to bringing up someone from nothing to best.

II. PROPOSED SYSTEM

With an increasing demand of English in the present world, it has become mandatory for everyone to be fluent in English in order to avoid any sort of communication issues throughout the globe, since English is the universal language for all. Our system aims to overcome the cumbersome and inconvenient way of acquiring proper communication skills in English language. The software's main objective is to enhance the user's way of speech through correctness of pronunciation by following the English Phonetics.

The system allows one to learn, judge and recognize their potential in English language. It includes various modules which ensure to provide a duo of fun time along with learning experience. This system is based on smart intelligence, voice processing and speech recognition. The combination of all, aims to provide possibly the best software, which will increase one's communication skill to a higher level, if inculcated efficiently.

III. BENEFITS OF THE PROPOSED SYSTEM

Every aspect of the system is designed, considering the modern methods of implication and implementations to make certain that learning is at ease and there is nothing difficult. Some of the distinctive features of the system are:

3.1 Cryptography

The user account consists of various field details which includes the result of progression in each dictionary. Cryptography is imbibed to make sure that the changes in the progress are acquired through genuine learning and not through fake updates.

3.2 Smart Recognition

The voice recording process is achieved through different recognition modus operandi like smart recognition or minimal recognition. Through smart recognition, the background sound can be filtered and only the efficient matching of words is processed.

3.3 MCQ and Tests

Learning overnight for an exam and forgetting everything just after it finishes is never a suggested way of learning. In the same way, learning new words or increasing the vocabulary, requires evaluation at every levels or intervals to ensure efficiency and firmness.

3.4 Confluence of separate modules

Merging the two different modules i.e. speech recognition module and text to speech module together onto the same platform, facilitating the user to acquire the benefits of both simultaneously.

IV. METHODOLOGY

As an initial attempt a simple model which converts speech to text and vice versa is developed using visual studio, SQLite and Microsoft speech recognition engine. The diagrammatic representation of this architecture is given in Fig. 1. The entire flow of the system is given in Fig. 2.

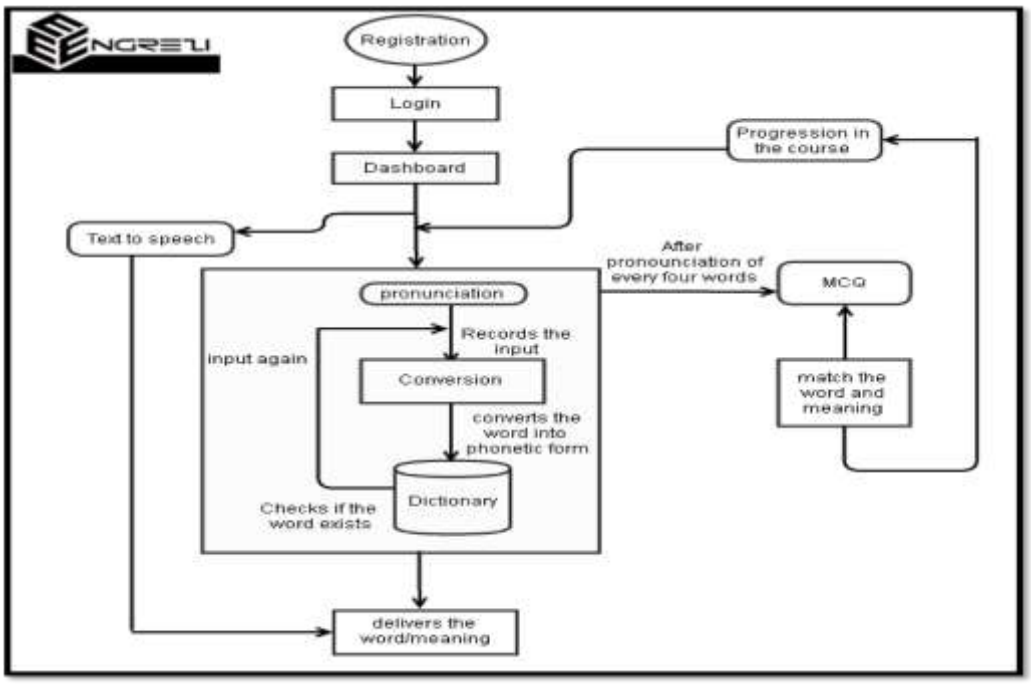

Fig. 1: Architecture diagram

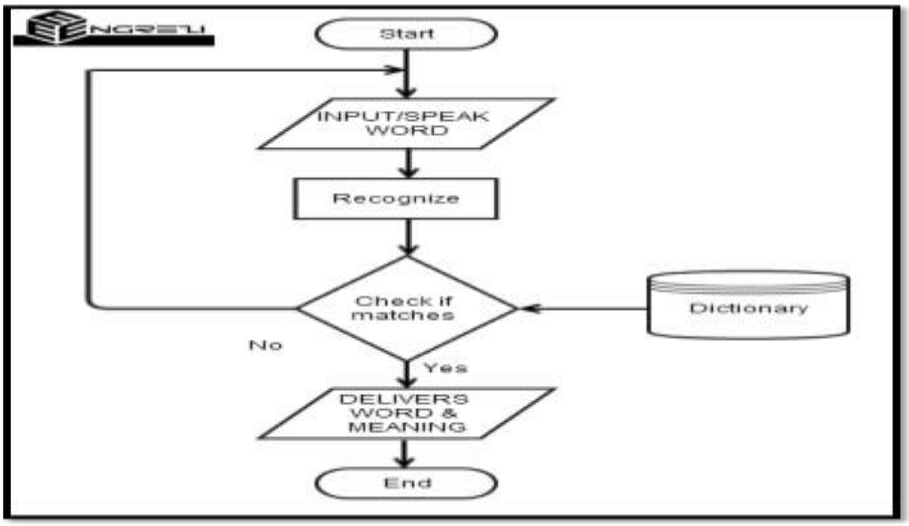

Fig. 2: Flow Diagram

V. LITERATURE REVIEW

In [1]Yee-Ling Lu , Man-Wai and Wan-Chi Siu explains about text-to-phoneme conversion by using recurrent neural networks trained with the real time recurrent learning (RTRL) algorithm. As recurrent neural networks deal well with spatial temporal problems, they are proposed to tackle the problem of converting

English text streams into their corresponding phonetic transcriptions.They We found that, due to the high computational complexity, the original RTRL algorithm takes a long time to finish the learning and proposed a fast RTRL algorithm (FRTRL), with a lower computational complexity which helps in fast learning.

In [2]Decadt, Jacques,Daelemans, Walter and Wambacq describes a method to enhance the readability of the textual output in a large vocabulary continuous speech recognition system when out-of-vocabulary words occur. The basic idea is to replace uncertain words in the transcriptions with a phoneme recognition result that is post-processed using a phoneme-to-grapheme converter. This technique uses machine learning concepts.

In [3]Penagarikano, M.; Bordel, G explains a technique to perform the speech to text conversion conversion as well as an experimental test carried out over a task oriented Spanish corpus are reported. They have concluded that the whole speech-to-text system neatly outperforms the word-constrained baseline system.

Martinez, M.; Quilis, A.; Bernstein, J have done a research aiming to develop a text-to-speech converter (TSC) for Spanish, that accepts a continuous source of alphanumeric characters (up to 250 words per minute) and produces good quality, natural Spanish output, is described. Four sets of problems are considered in this work: the hard-ware structure adopted for real time operation; the complex control software needed to handle the orthographic input and linguistic programs; the linguistic processing rules, and the parameterization of the Spanish language matched to a TSC. Emphasis is made on the problems of adapting a general hardware structure to a specific language[4].

In [5]Sultana, S.; Akhand, M. A H; Das, P.K.; Hafizur Rahman, M.M. investigate Speech-to-Text (STT) conversion using SAPI for Bangla language .They says that experimental study was carried out for the technique on an article from a news paper and the recognition rate was approximately 78% on an average. Although achieved performance is promising for STT related studies, they identified several elements to improve the performance and might give better accuracy and assures that the theme of this study will also be helpful for other languages for Speech-to-Text conversion and similar tasks.

Moulines, E., in his paper "Text-to-speech algorithms based on FFT synthesis," present FFT synthesis algorithms for a French text-to-speech system based on diphone concatenation. FFT synthesis techniques are capable of producing high quality prosodic modifications of natural speech. Several approaches are presented to reduce the distortions due to diphone concatenation[6].

VI. CONCLUSION AND FUTURE SCOPE

This project aims to provide an easy platform to learn and master the English language with modern ways of technology. It includes the correctness of spelling and meaning with end results of achieving excellence in pronunciation. In future we are planning to improve the pronunciation i.e. sound accuracy by incorporating appropriate filtering techniques. Comparative study of the existing TTS and STT algorithms are performed and work has to be done to improve the performance & improve the quality of the output.

The project has been planned to be desgined in a way that it is the complete course learning process for the betterment of pronunciation of the users struggling to achieve at convenience.

REFERENCES

[1]. Yee-Ling Lu; Mak, Man-Wai; Wan-Chi Siu,, "Application of a fast real time recurrent learning algorithm to text-to-phoneme conversion," Neural Networks, 1995. Proceedings., IEEE International Conference on , vol.5, no., pp.2853,2857 vol.5, Nov/Dec 1995.

[2]. Decadt, Bart; Duchateau, Jacques; Daelemans, Walter; Wambacq, P., "Phoneme-to-grapheme conversion for out-of-vocabulary words in large vocabulary speech recognition," Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on , vol., no., pp.413,416, 2001.

[3]. Penagarikano, M.; Bordel, G., "Speech-to-text translation by a non-word lexical unit based system," Signal Processing and Its Applications, 1999. ISSPA '99. Proceedings of the Fifth International Symposium on , vol.1, no., pp.111,114 vol.1, 1999.

[4]. Olabe, J. C.; Santos, A.; Martinez, R.; Munoz, E.; Martinez, M.; Quilis, A.; Bernstein, J., "Real time text-to-speech conversion system for spanish," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84. , vol.9, no., pp.85,87, Mar 1984.

[5]. Sultana, S.; Akhand, M. A H; Das, P.K.; Hafizur Rahman, M.M., "Bangla Speech-to-Text conversion using SAPI," Computer and Communication Engineering (ICCCE), 2012 International Conference on , vol., no., pp.385,390, 3-5 July 2012.

[6]. F.; Moulines, E., "Text-to-speech algorithms based on FFT synthesis," Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on , vol., no., pp.667,670 vol.1, 11-14 Apr 1988.