ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 5, Issue 1, January 2016

English Text to Speech Conversion using Indian Pronunciation

Kaladharan N

Abstract -The objective of this paper is to change the text into speech. Speech synthesis is the artificial method of human speech. Text supervision and speech creation are the two main parts of the text to speech system. Realness and fluency are mainly significant, while making a text to speech system. Database creation, character recognition and text to speech changing are the important parts in text to speech analysis.

Key words- Text to Speech system, Speech synthesis, Synthesis tool, MOS

I.INTRODUCTION

Speech is the most frequently and normally used method of communication between persons. The artificial making of human speech by machines or computers is named Speech Synthesis. The word 'Synthesis' is defined by the words as 'fusion/mixing'. Text to speech process is also called as Speech synthesis [1]. Generally, the text to speech system is to use to convert an arbitrary input text to corresponding natural sounding speech in a more intelligible way [2]. Mostly, a text by means of sentences is composed of group of words, whereas words are grouping of alphabets arranged in an expressive manner. Naturalness and intelligibility are the essential quality of the text to speech synthesis. Naturalness defines the output speech sounds similar to human speech and intelligibility is which the output speech sound is understood. Text processing and speech generation are the two main components of a text to speech system [3][4]. The task of the text processing part is to generate suitable sequence of phonemic items and speech generation component is to synthesize the acoustic waveform of speech or by the speech database. The TTS system can be used to read text from emails, epapers, Interactive voice responses, Short message services, web pages, news, articles, blogs, talking books and toys, games, man machine communications etc. [5].

Kaladharan N, *Assistant Professor, Department of Electrical and engineering, Annamalai University, India.*

This paper presents a system to design a Text to Speech conversion. This paper is structured as follows section-II techniques of speech synthesis, sectionIII-proposed system, section IV-Implementation, section V-Results, section VI-Performance measure and section VII-Conclusion.

II. TECHNIQUES OF SPEECH SYNTHESIS

There are different ways to perform speech synthesis:

Concatenative synthesis is based on the concatenation of slices of recorded speech. This speech synthesis uses phones, di phones, syllables, words or sentences as basic speech units. It involves two phases, viz., the offline and online phase. Offline phase includes preprocessing, segmentation and pitch design. Online phase includes text examination and synthesis [6]. Formant synthesis is also known as rule based synthesis, creates the acoustic speech data fully through procedures on the acoustic compares of the several speech sounds. Formant synthesized speech is reliably clear, even at very high speeds. Diphone synthesis is most popular way used for creating a synthetic speech from recordings or samples of a particular person. In Diphone synthesis, the strength of speech depends on expression or sentence and the model used for prosody. It uses a small speech database. Domain specific synthesis concatenates prerecorded words and phrases to generate whole utterances. It is used in applications like transit schedule broadcasts or weather reports, talking clocks and calculators. This is very simple to implement .Unit selection synthesis is the dominant synthesis technique in text to speech [7]. Unit selection synthesis uses huge databases of recorded speech.

III. PROPOSEDSYSTEM

The block diagram of the proposed system is shown in Fig.1.

3.1. Phonemes

In any language, a phoneme is a single part of sound that has a valuable meaning. There are 44 phonemes in English, each one signifying an altered sound a person can make. Since there are only 26

letters in the alphabet, sometimes letter groupings need to be used to make a phoneme. A letter can also characterize different phonemes.

3.2. Wave file

A Wave file is an audio file format, made by Microsoft. This is the standard computer audio file format. A Wave file is identified by a file name extension of WAV (.wav). In addition to the uncompressed raw audio data, the Wave file format stocks information about the file's number of tracks mono or stereo, sample rate, and bit depth etc.

3.3. Text or .Txt

Text is a human readable sequence of characters and the words they form that can be encoded into computer readable formats. TXT stands for Text. TXT is a file extension for a text file, used by a variety of text editors.

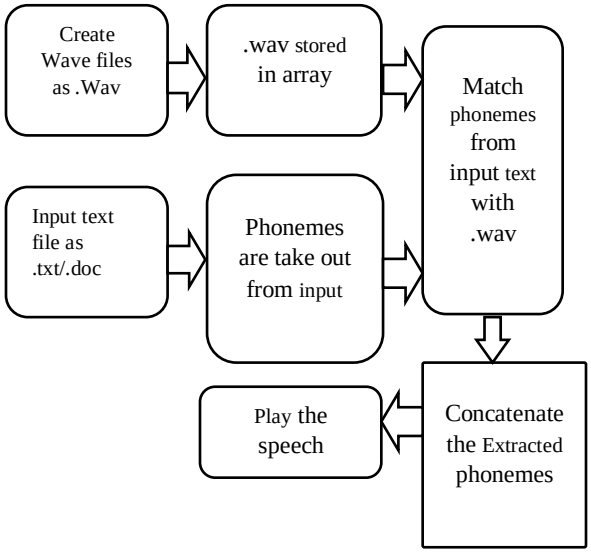


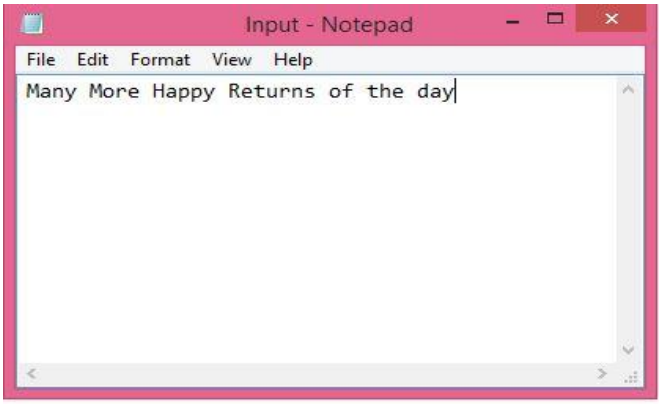
Fig. 1. Block diagram of planned system

The voice which is recorded manually contains some gap and it takes more memory space to save and also it makes the trouble of speech sample is to listen were unkind and unusual. So the removal of the Pause or gap is very essential for hearing a correct speech.

IV. IMPLEMENTATION

Create the wave files of the related text by using the microphone and stored in an array as .Wav format. Create the text file, which contains the desired sentence or words as input and stored as .txt or .doc file. Extract the phonemes from the input text file and compare with the wave file. Then concatenate the phonemes and play wave file.

The database was produced using open source recording software's and saved in the folder in desktop. Record all the words of adictionary, the database memory also bigger. Hence choosing sound element with appropriate length is the important criteria to receive theword is to be natural andunderstandable when synthesized. New other database set was created and tested of English language in Indian pronunciation using the same method and the input recording was sentences.



V. RESULTS

Fig.2. Text input from notepad
Fig.3.Speech Output

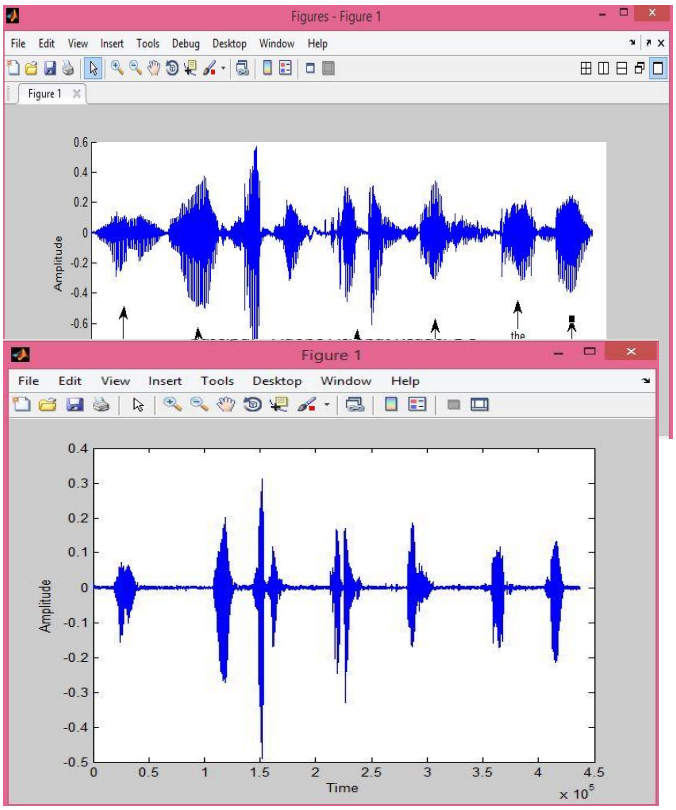


Fig.4.Pause removed speech output

The fig.2 text file from notepad is given as the input of the TTS system. This file contains the text

'*Many more happy returns of the day*' is to converting into speech. The output speech of the given sentences is shown in fig.3. And the removed gaps of the sentences should be concatenated and played is shown fig.4

VI. PERFORMANCE MEASURE

Listening tests is supported to evaluate the quality of the system. By subjective tests method, 15 individual listeners listen the audio quality and they rate the given input sentences by MOS. Mean Opinion Score is composed of five different scores of subjective quality following by the table.1. The Mean opinion score of genuineness and clarity is show in the fig.5 and fig.6.

Mean opinion value	Class
5	Perfect
4	Nice
3	Worthy
2	Reasonable
1	Inferior

Tab 1: Mean opinion score approvals

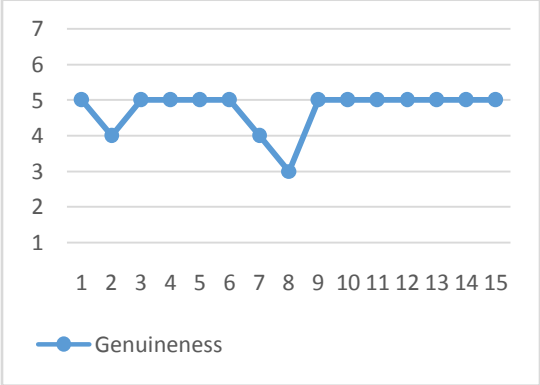


Fig. 5: Mean Opinion Score for Genuineness

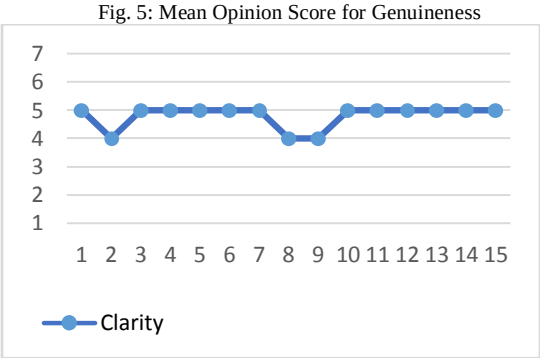


Fig.6 Mean Opinion Score for Clarity

VIICONCLUSION

The proposed system gives a better result for text to speech conversion. After performing individual listening test satisfactory results were obtained. This system can only synthesize the sentences that are present in the document. Five classes were perfect, nice, worthy, reasonable, and inferior for scoring Mean Opinion Score (MOS) was used. Each listener was asked to provide the score by individually. Thus this system very easy and efficient to implement. In future, further improvement can be done on the quality of synthesizer for expression and emotions in various situations and give a real time applications.

REFERENCES

[1] .Kaladharan N, "An English Text to Speech Conversion System",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 10, October-2015

[2].Shreekanth.Ta, et.al "Duration Modelling Using Neural Networks for Hindi TTS System Considering Position of Syllable in a Word"International Conference on Information and Communication Technologies.

[3]. Suhas R. Mache,et.al, "Review on Text-To-Speech Synthesizer, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 8, August 2015

[4]. Lakshmi Sahu, "Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion, International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012

[5]. J.O.Onaolapo,et.al, "A Simplified Overview of Text-To-Speech Synthesis, Proceedings of the World Congress on Engineering 2014 Vol I,

[6] .Helal Uddin Mullah ,"A Comparative Study of Different Text-to-Speech Synthesis Techniques",International Journal of Scientific & Engineering Research, Volume 6, Issue 6, June-2015

[7].A.Indumathi, et, al, "Survey On Speech Synthesis", Signal Processing: An International Journal (SPIJ), Volume (6): Issue (5): 2012