



International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijirccce.com

Vol. 5, Issue 12, December 2017

Text to Speech Conversion for Vocally Disable

Atharva Chogle¹, Shweta Baviskar²

Bachelor of Engineering, Dept. of Computer Engineering, RGIT, Mumbai, Maharashtra, India¹

Bachelor of Engineering, Dept. of Electronics and Telecommunication Engineering, RGIT, Mumbai, Maharashtra, India²

ABSTRACT:Text to speech (TTS) synthesis is the automatic conversion of text into speech. Generally, TTS system consists of two phases. The first is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation. The second one is the generation of speech waveforms. In this TTS system, text to phoneme conversion depends on dictionary based approach to get the exact phonetic transcription. Speech synthesis such as domain specific, phoneme based synthesis and unit selection synthesis are used for concatenating speech. For numerical text to speech system, domain specific synthesis is applied. In phoneme based synthesis, the input text is considered as word to produce sound. For input sentence. Unit selection speech synthesis is applied. This TTS system is mainly used for visual impairments and handicapped people [1].

KEYWORDS: Text to Speech (TTS), transcribed, phoneme, linguistic

I. INTRODUCTION

The text-to-speech (TTS) synthesis is to convert an arbitrary input text into intelligible and natural sounding speech. TTS system includes mainly two parts: natural language processing and digital signal processing. The general block diagram of TTS system is shown in figure 1. Natural language processing contains three steps. They are text analysis, phonetic analysis and prosodic analysis. The text analysis includes segmentation, text normalization, and part of speech (POS) tagger. Phonetic conversion is to assign phonetic transcription to each word. There are two approaches in phonetic conversion. They are rule based and dictionary based approaches. Rule based is applied for known words whereas dictionary based is used for unknown words. Prosodic analysis is to determine intonation, amplitude and duration modelling of speech. It describes speaker's emotion [1].

II. METHODOLOGY

Text to speech system has two parts namely natural language processing and speech synthesis (digital signal processing).

(1) Natural Language Processing (NLP):- NLP produces phonetic transcription together with prosodic feature of the input text. In this TTS system, NLP comprises of three main components such as text analysis, phonetic conversion and prosodic phrasing [1].

(2) Text analysis: - In this TTS system, the input sentence is segmented into token. After tokenization, each word is determined as part of Speech tagging [2].

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

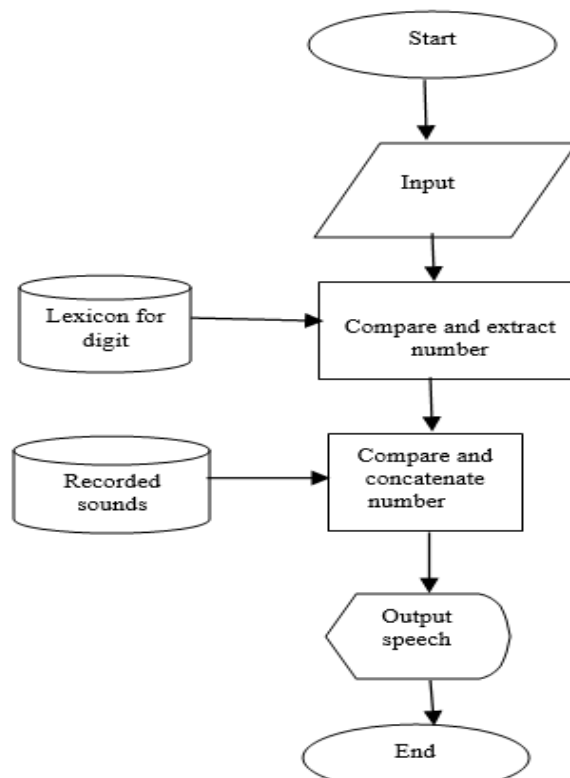


Figure 1.Flowchart on Text to speech (TTS) synthesis

A. System Objectives:-

The basic objective of this system is to convert a text signal into a audio signal in different voices. It allows to access the content more easily in audio format.

B. Proposed Systems:-

People who have learning disabilities can speak the language and due to the listening skills they can understand the content easily.

C. System details:-

Here the question arises whether the System can convert the text to speech .The answer is absolutely yes as the method which is used is the Function of Text to Speech.

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

PHONEME BASED SELECTION:-

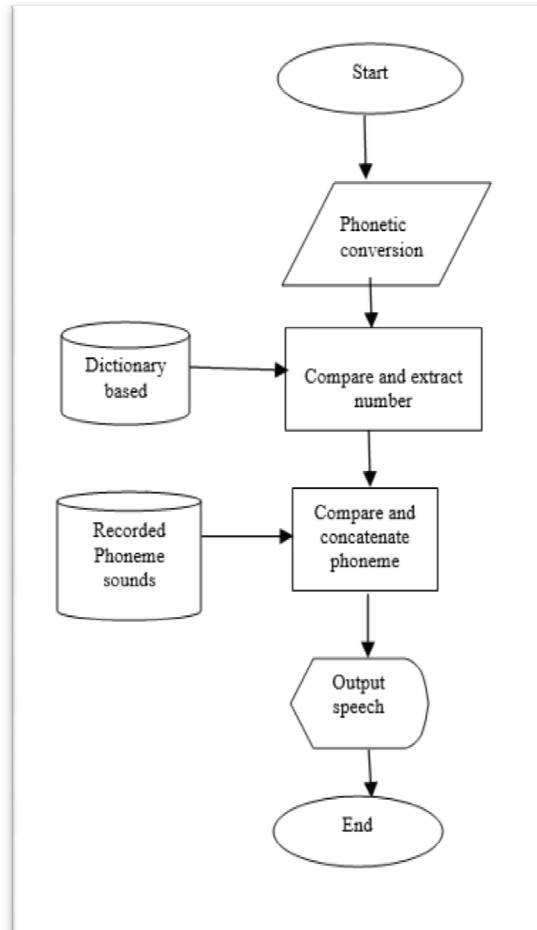


FIGURE 2.FLOWCHART ON PHONEME BASED TTS SYNTHESIS

III. DESCRIPTION

Software implementation is based on Matlab programming language. In domain specific synthesis, the input number (one or more digits) can be pronounced speech easily and quickly. The output speech is natural and intelligible like human speech. But the domain specific synthesis is not general purpose .For the combination of words; it must be preprogrammed to synthesise speech. Authors and Affiliations.

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

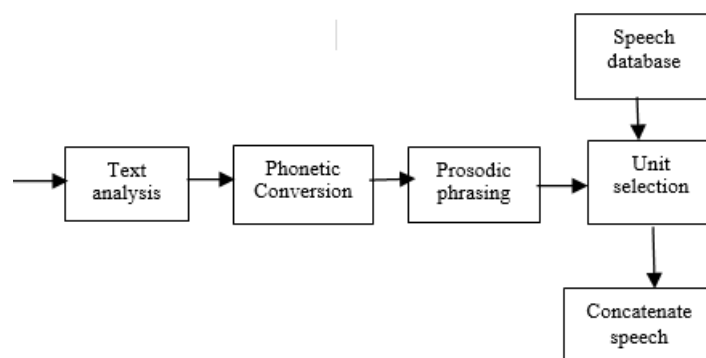


Figure 3. Block diagram on TTS synthesis

In speech synthesis, speech is recorded at sampling rate of 16kHz. Then these recorded speeches are segmented phonetically in discrete time domain. So the segmented units are stored according to their sample values. There are two approaches in speech segmentation. They are hand-labelling and automatic speech segmentation. In this TTS system, hand-labelling is applied because database is small. But it is time consuming and has little errors in this method. Unit Selection of algorithm which selects the best acoustic units which match the target linguistic features. Then these units are concatenated to produce speech [1].

IV. RESULTS

The Matlab program to convert text to speech works efficiently with proper voice modulation. Here the voice modification can be effectively done from a man to woman or vice versa. With the help of modification of speech assistant present in control panel e.g. David, Hazel, etc. Here a vocally disabled person can use the program effectively with the help of keyboard and it is possible for him/her to talk even on cell phones or to a person (physically).

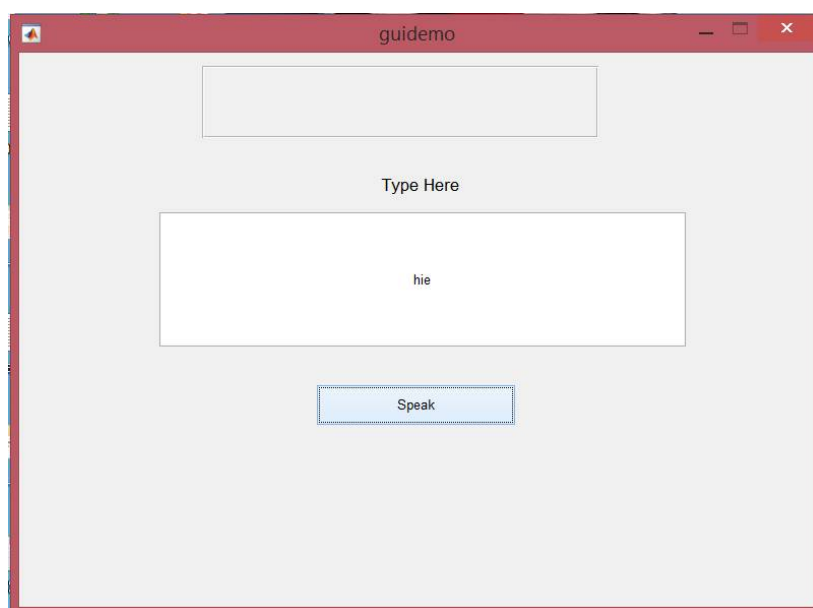


Figure 4. Output of Text to speech

International Journal of Innovative Research in Computer and Communication Engineering

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

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

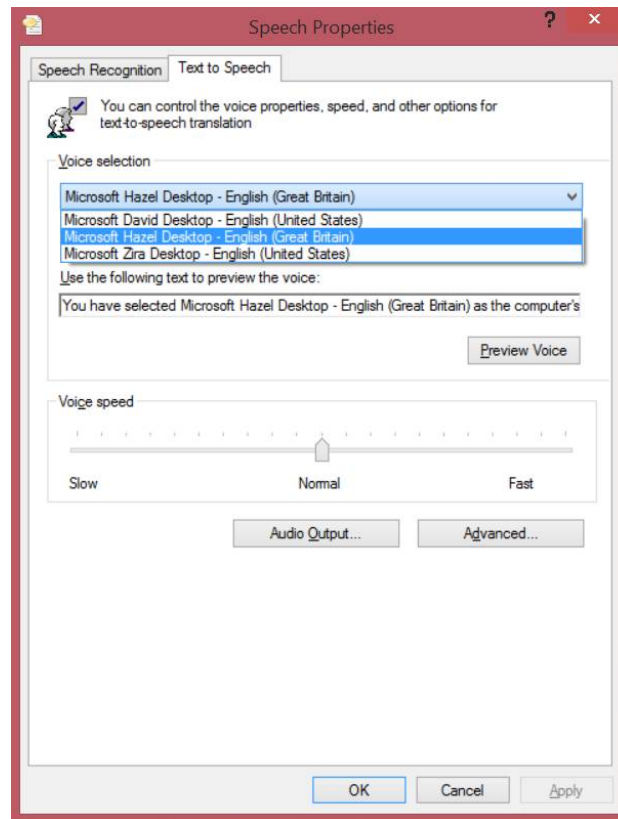


Figure 5. Process of changing voice

V. CONCLUSION

In this paper, text to speech system is developed for numbers, words and sentence. For numbers (one or more digits), the output speech is natural and pleasant to listen. It is necessary to remove delay in speech when speech waveforms are done concatenation. So, domain specific synthesis can smoothly produce speech. But the output speech of words are discontinuities between transitions of phoneme. For two or more syllable words, syllabification method is more appropriate. The sound quality is intelligibility. Thus, domain specific and phoneme based synthesis are very easy and efficient to implement unlike other methods which involve many complex algorithms. But in unit selection synthesis, the implementation is not easy as these two methods. The output sentence of speech has little glitch. However, the speech output is better than phoneme based synthesis [1].

VI. ACKNOWLEDGEMENT

This research was supported by Mrs **Jyoti deshmukh**(Assistant professor) , I sincerely Thank her for providing me an opportunity to do the project research on text to speech synthesis using Data mining at Rajiv Gandhi Institute of Technology (R.G.I.T.), Andheri, Mumbai. Text To Speech Conversion for Visually Impaired People,”

REFERENCES

- [1]International Journal of Scientific & Technology Research Volume 4, ISSUE 07, JULY 201 5 ISSN 2277-8616
- [2]International Journal of Applied Information Systems (IJ AIS) – ISSN: 2249-0868 Foundation of Computer Science CS, New York, USA Volume 7– No. 2, April 2014 – www.ijais.orgItunuoluwa Isewon, “Design and Implementation of Text To Speech Conversion for Visually Impaired People,”