International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-6, Issue-8)

Research Article | **August 2017**

Error Free Punjabi Text to Speech Generation System based on Phonemes

Tejinder Kaur, Charanjiv Singh
Department of Computer Engineering, Punjabi University, Patiala,
Punjab, India

Abstract-

Text-to-speech (TTS) is the generation ofsynthesized speech from text.Language is the ability to express one'sthoughts by means of a set of signs (text), gestures,and sounds. It is a distinctive feature of humanbeings, who are the only creatures to use such asystem. Speech is the oldest means of communicationbetween people and it is also the most widely used.'Speech synthesis' also called 'Text to speechsynthesis' is the artificial production ofhuman speech. A computer system used for thispurpose is called a speech synthesizer and can beimplemented in software. A text-to-speech(TTS) system converts text to speech.The proposed Enhanced Transcriptions Method is developed using Microsoft Visual Studio in VB.Net Language. Firstly word indexing is performed for the predefined words then corresponding speech signal is detected and errors in words are calculated using Euclidean distance. The results of the proposed work shows that Enhanced Transcriptions Method has more accuracy 89% as compared to previous Transcriptions Method 79%. The value of specificity for proposed method is 0.89 and for previous method is 0.79.

Keywords- Enhanced Transcriptions Method, Phonemes, Text to Speech, Euclidean distance, Speech Synthesis.

I. INTRODUCTION

A. Text To Speech

Speech synthesis is the synthetic manufacturing of human speech. A pc machine used for this purpose is called a speech laptop or speech synthesizer, and can be carried out in software or hardware products. A textual content-to-speech (TTS) gadget converts ordinary language textual content into speech; other structures render symbolic linguistic representations like phonetic transcriptions into speech. Synthesized speech can be created by means of concatenating pieces of recorded speech which can be stored in a database. Systems fluctuate within the length of the saved speech gadgets; a machine that stores telephones or diaphones provides the most important output variety however may lack readability. For unique usage domains, the garage of complete words or sentences lets in for terrific output. Alternatively, a synthesizer can include a model of the vocal tract and different human voice characteristics to create a very "synthetic" voice output. The first-class of a speech synthesizer is judged with the aid of its similarity to the human voice and by way of its capacity to be understood certainly. Intelligible text-to-speech software lets in people with visual impairments or analyzing disabilities to listen to written works on a domestic pc. Many pc running systems have protected speech synthesizers because the early 1990s.

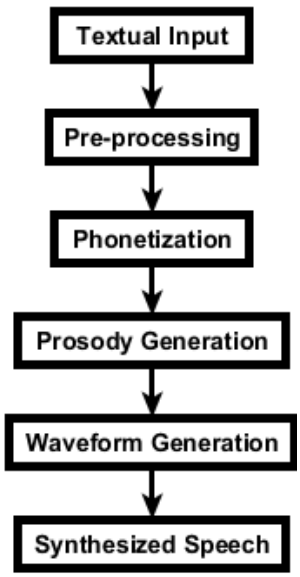

Fig. 1: Overview of TTS system

Textual content-to-speech gadget (or "engine") is composed of two parts: the front-end and a back-cease. The front-stop has two fundamental duties. First, it converts uncooked textual content containing symbols like numbers and abbreviations into the equivalent of written-out phrases. This method is frequently called textual content normalization, pre-processing, or tokenization. The front-stop then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic devices, like phrases, clauses, and sentences. The method of assigning phonetic transcriptions to words is referred to as textual content-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody statistics together make up the symbolic linguistic illustration this is output by using the front-cease. The returned-give up—frequently called the synthesizer—then converts the symbolic linguistic representation into sound. In positive systems, this part consists of the computation of the target prosody (pitch contour, phoneme intervals), that is then imposed on the output speech [8].

B. Tokenization

Tokenization is the technique of replacing sensitive data with specific identity symbols that maintain all the important information about the records without compromising its security. Tokenization, which seeks to decrease the amount of records a commercial enterprise desires to preserve accessible, has end up a popular manner for small and mid-sized businesses to reinforce the security of credit card and e-commerce transactions even as minimizing the cost and complexity of compliance with industry standards and government rules [10].

Payment card industry (PCI) requirements do no longer permit credit card numbers to be stored on a retailer's factor-of-sale (POS) terminal or in its databases after a transaction. To be PCI compliant, traders need to installation expensive give up-to-quit encryption systems or outsource their payment processing to a provider who offers a "tokenization option." The carrier provider handles the issuance of the token value and bears the obligation for keeping the cardholder statistics locked down.

In this kind of state of affairs, the provider company problems the merchant a driving force for the POS gadget that converts credit score card numbers into randomly-generated values (tokens). Since the token is not a primary account wide variety (PAN), it cannot be used out of doors the context of a particular transaction with that specific service provider. In a credit card transaction, for instance, the token usually carries handiest the last 4 digits of the real card variety. The relaxation of the token includes alphanumeric characters that represent cardholder information and statistics particular to the transaction underway.

Tokenization makes it extra hard for hackers to benefit get right of entry to cardholder facts, as compared with older systems wherein credit card numbers had been stored in databases and exchanged freely over networks. Tokenization era can, in idea, be used with sensitive facts of all kinds consisting of bank transactions, clinical statistics, crook statistics, car motive force records, mortgage programs, inventory trading and voter registration [10].

C. Transcription

Transcription is step one of gene expression, in which a selected segment of DNA is copied into RNA (especially mRNA) through the enzyme RNA polymerase. Both DNA and RNA are nucleic acids, which use base pairs of nucleotides as a complementary language. During transcription, a DNA collection is study by an RNA polymerase, which produces a complementary, antiparallel RNA strand referred to as a primary transcript [11].

Transcription proceeds in the following general steps:

1. RNA polymerase, together with one or greater well-known transcription elements, binds to promoter DNA.
2. RNA polymerase creates a transcription bubble, which separates the 2 strands of the DNA helix. This is done with the aid of breaking the hydrogen bonds among complementary DNA nucleotides.
3. RNA polymerase provides RNA nucleotides (that are complementary to the nucleotides of 1 DNA strand).
4. RNA sugar-phosphate spine forms with assistance from RNA polymerase to shape an RNA strand.
5. Hydrogen bonds of the RNA–DNA helix damage, freeing the newly synthesized RNA strand.
6. If the cell has a nucleus, the RNA may be similarly processed. This might also consist of polyadenylation, capping, and splicing.
7. The RNA may continue to be inside the nucleus or go out to the cytoplasm through the nuclear pore complex.

The stretch of DNA transcribed into an RNA molecule is called a transcription unit and encodes at least one gene. If the gene encodes a protein, the transcription produces messenger RNA (mRNA); the mRNA, in turn, serves as a template for the protein's synthesis through translation. Alternatively, the transcribed gene may encode for non-coding RNA (including microRNA), ribosomal RNA (rRNA), switch RNA (tRNA), or other enzymatic RNA molecules known as ribozymes. Overall, RNA helps synthesize, adjust, and manner proteins; it consequently performs a fundamental function in acting capabilities within a cell.

In virology, the term will also be used whilst referring to mRNA synthesis from an RNA molecule (i.e., RNA replication). For example, the genome of a negative-feel unmarried-stranded RNA (ssRNA -) virus can be template for an effective-sense unmarried-stranded RNA (ssRNA +). This is because the wonderful-experience strand carries the records needed to translate the viral proteins for viral replication afterwards. This technique is catalyzed with the aid of viral RNA replicas [11].

II. LITERATURE REVIEW

Arlo Faria et.al in [1] described a text-to-speech utility for a spread of Brazilian Portuguese. After offering the language's phonetic attributes, the orthographic gadget is examined and shown to be a function that maps letters to those sounds.

Given the orthography's phonological regularity, it is simple to implement the textual analysis part of a speech synthesis machine, as I exhibit with some simple Perl code.
Ruvan Weerasinghe et.al in [2] brings collectively the improvement of the first Text-to- Speech (TTS) system for Sinhala the use of the Festival framework and practical programs of it. Construction of a diphone database and implementation of the herbal language processing modules are defined. The paper additionally presents the improvement method of direct Sinhala Unicode textual content input by using rewriting letter-to-sound policies in Festival's context touchy rule format and the implementation of Sinhala syllabification algorithm. A Modified Rhyme Test (MRT) changed into conducted to evaluate the intelligibility of the synthesized speech and yielded a rating of seventy one.5% for the TTS machine described.
Kalika Bali et.al in [3] described in detail a Grapheme-to-Phoneme (G2P) converter required for the improvement of an awesome satisfactory Hindi Text-to-Speech (TTS) machine. The Festival framework is chosen for growing the Hindi TTS device. Since Festival does now not provide complete language processing assist precise to numerous languages, it needs to be augmented to facilitate the development of TTS structures in positive new languages. Because of this, a general G2P converter has been evolved. In the custom designed Hindi G2P converter, we've handled schwa deletion and compound phrase extraction. In the experiments done to check the Hindi G2P on a textual content phase of 3485 phrases, 97.67% phrase phonetisation accuracy is acquired. This Hindi G2P has been used for phonetising huge textual content corpora which in turn are used in designing an inventory of coverage of the phonetically valid diaphones the usage of simplest zero.3% of the complete textual content corpora.
Sangramsing et.al in [4] Describe in detail a Grapheme-to-Phoneme (G2P) converter required for the improvement of a great pleasant Marathi Text-to-Speech (TTS) machine. The Festival and Fest ox framework is chosen for developing the Marathi TTS gadget. Since Festival does now not provide complete language processing help specie to various languages, it needs to be augmented to facilitate the improvement of TTS systems in positive new languages. Because of this, a normal G2P converter has been developed. In the customized Marathi G2P converter, we've got treated schwa deletion and compound word extraction. In the experiments achieved to test the Marathi G2P on a textual content section of 2485 words, ninety one.47% word phonetisation accuracy is received. This Marathi G2P has been used for phonetising massive textual content corpora which in flip are utilized in designing an stock of phonetically wealthy sentences. The sentences ensured an excellent insurance of the phonetically legitimate di-telephones the use of only 1.Three% of the complete text corpora.
Deepa S.R. et.al in [5] addressed the trouble of Hindi compound word splitting and its relevance to developing an amazing quality phonetizer for Hindi Speech Synthesis. The constituents of a Hindi compound phrase aren't separated by way of space or hyphen. Hence, maximum of the prevailing compound splitting algorithms cannot be implemented to Hindi. We endorse a brand new technique for automated extraction of compound words from Hindi corpus. Preliminary exams carried out at the algorithm have proven a breakup fee of 92 to 96% of the input compound words. Of these splits, around 83 to 87% are accurate splits. A few adjustments were counseled, with a purpose to enhance the accuracy of the splits. Finally, we examine an development of 1.6% in Hindi Grapheme-to-Phoneme (G2P) conversion because of the use of a phonetized compound phrase lexicon, created by way of the above approach.
Sangramsing et.al in [6] provided the technique toward changing textual intent to speech the use of new technique. The text to speech conversion device permits user to enter textual content in Marathi and as output it gets sound. The paper provides the steps accompanied for changing textual content to speech for Marathi language and the algorithm used for it. The consciousness of this paper is primarily based at the tokenization procedure and the orthographic illustration of the text that suggests the mapping of letter to sound using the description of language's phonetics. Here the primary focus is on the text to IPA transcription concept. It is in reality, a system that interprets textual content to IPA transcription that's the primary stage for textual content to speech conversion. The entire manner for converting text to speech involves an extraordinary deal of time as it's now not an easy assignment and requires efforts.
Soumya Priyadarsini Panda et.al in [7] offered a top level view of the TTS synthesis technology in conjunction with details of the stages concerned. The thrust has been given to explore the usefulness of this method in designing a TTS system for Indian languages. This paper also specializes in a number of the open studies troubles wherein paintings on this region may additionally similarly be carried out.

III. PROBLEM FORMULATION

In recent years, the use of computers in speech synthesis and speech recognition has become an important area of study among speech and computer scientists. The primary motivations are to provide users with a friendly vocal interface with the computer and to allow people with certain handicaps (such as blindness) to use the computer.
Speech is the primary means of communication between people. Punjabi language is being spoken by about 104 million peoples in India, Pakistan and other countries with Punjabi migrants. The language is being written in Gurmukhi script in Indian Punjab, whereas in Shahmukhi script in Pakistani Punjab. Every word in Punjabi or any language has its well defined boundaries. sometimes there are other problems that may come by using word as a speech unit. Each word has to be trained individually and there any sharing of parameters cannot be possible among words. so, it is important to have a large training set so that all words in vocabulary are adequately trained. When all conversions are performed from text to speech and it comes to pronunciation of word then various grammatical and pronunciation errors occur due to fact that a single word can be written in different ways. Therefore need of such system is felt which can detect the changes in original word and pronunciate it correctly as original word. We proposed to design and implement such system which can smartly convert text into speech and pronunciate the word correctly.

IV. PROPOSED RESEARCH METHODOLOGY

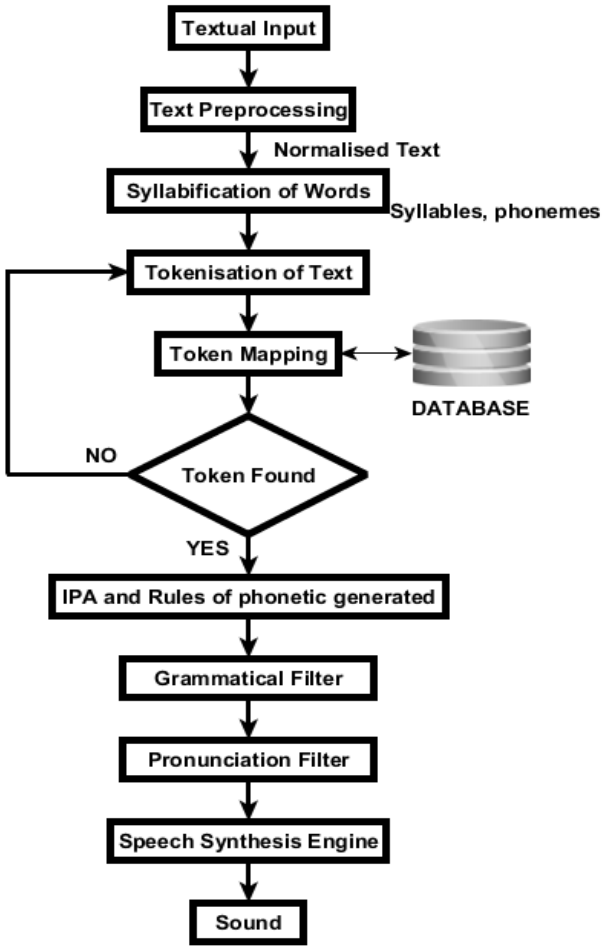

Fig. 2: Proposed Research Methodology

V. SIMULATION AND EXPERIMENTAL RESULTS

From the last decade, the quality of text-to-speech (TTS) [12] has been improved dramatically. In General, human-like clear sound can be generated from waveform-based speech synthesis but it necessarily has a huge amount of speech data with the control flexibility. At the same time, smooth sound can be generating from Artificial Neural Network(ANN) based speech synthesis with a small amount of speech data. Further, it has flexibility in controlling speaker individuality [13]. In ANN based speech synthesis, the system is developed by creating a library of phonemes, a library of phoneme audio files and a dictionary of words with their phoneme representation. The system generates takes a sentence, analyze each word and find out their corresponding phonemes. Then it concatenates all the phonemes from the phoneme audio library and then plays the audio which sounds like a speech of sentence. It also displays a waveform and spectrum of the generated speech sound. When all conversions are performed from text to speech and it comes to pronunciation of word then various grammatical and pronunciation errors occur due to fact that a single word can be written in different ways. Therefore need of such system is felt which can detect the changes in original word and pronunciate it correctly as original word. We proposed to design and implement such system which can smartly convert text into speech and pronunciate the words correctly.

```
0references
private void button1_Click(object sender, EventArgs e)
{
    reader.Dispose();
    if (textBox1.Text != "")
    {

        reader = new SpeechSynthesizer();
        reader.SpeakAsync(textBox1.Text);
        label2.Text = "SPEAKING";
        if (fname == "F1.txt" || fname == "F11.txt")
        {
            System.Media.SoundPlayer player = new System.Media.SoundPlayer("F1.wav");
            player.Play();
        }

        else if (fname == "F2.txt" || fname == "F12.txt")
        {
            System.Media.SoundPlayer player = new System.Media.SoundPlayer("F2.wav");
            player.Play();
```

Fig. 3: Code developed using Visual Studio

The code developed is shown in figure 3. The whole proposed system is designed using Microsoft Visual Studio and VB.net language is used for developing the code concept.

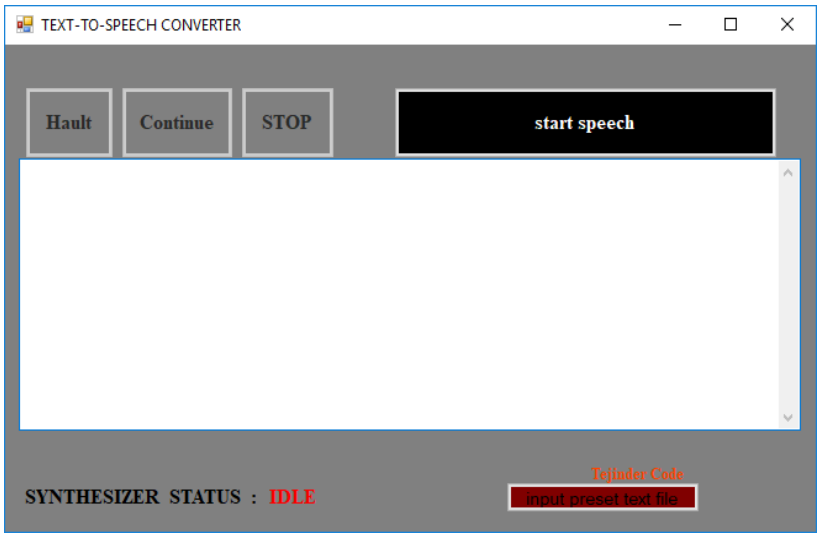

Fig. 4: Graphical User Interface (GUI)

The Graphical User Interface (GUI) of the proposed system is shown in figure 4. The GUI shows buttons like input preset text file, start speech, Halt, Continue and Stop.

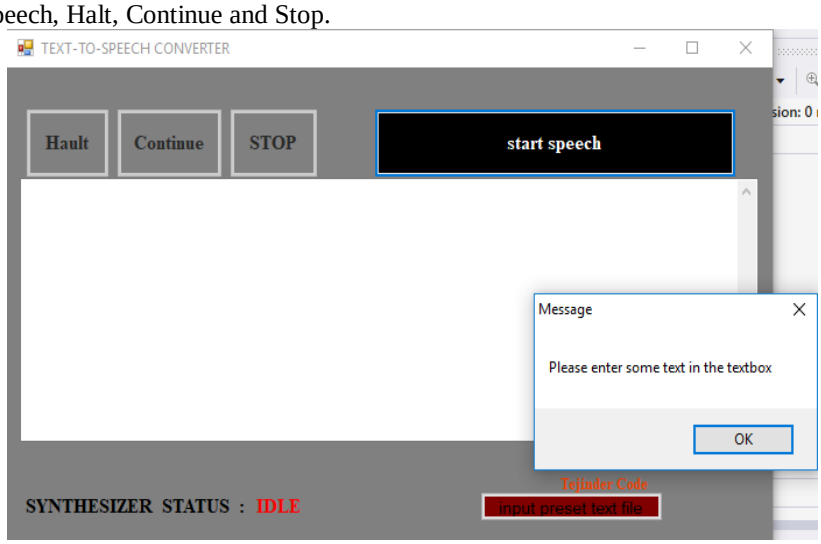

Fig. 5: Message Box

The Message box is shown in figure 5. This message box pop up, when start speech button is pressed before selecting the input text file. The message box displays message “Please enter some text in the text box”.

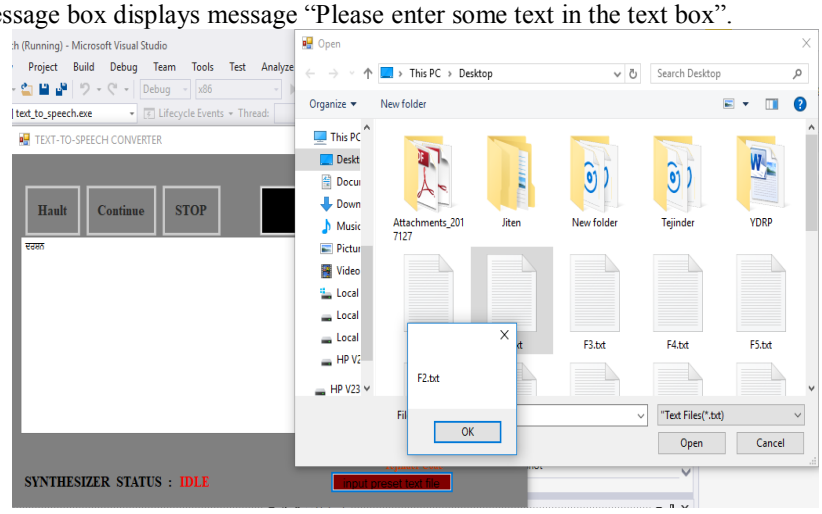

Fig. 6: Selecting input text file

The selection of input preset text file is shown in figure 6. A window popups when “input preset text file” button is pressed. When window popups we have to choose one text file as input. After input of text file is done message box pops up showing the name of selected text file.

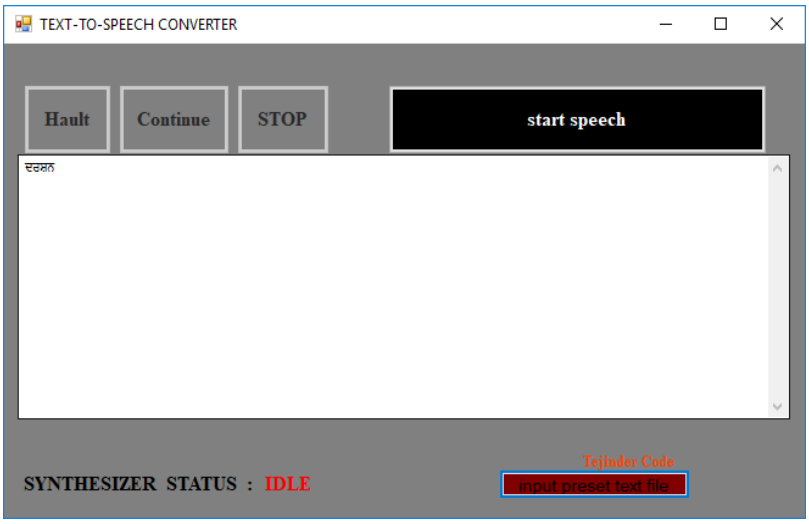

Fig. 7: Word from text file shown in text window

The figure 7 shows after the selection of input text file, the word stored in the text file is shown in text window. After that we can press "start speech" button and listen the speech.

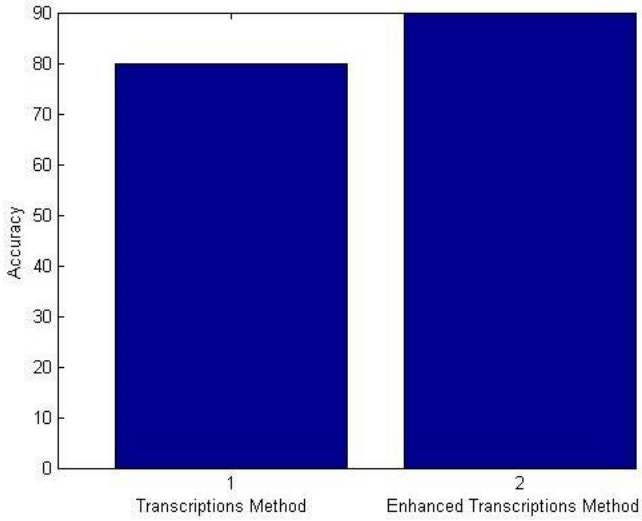

Fig. 8: Accuracy

The accuracy of the previous (Transcriptions Method) and proposed (Enhanced Transcriptions Method) is shown in figure 8. The accuracy of previous method is 79% and proposed method is 89%.

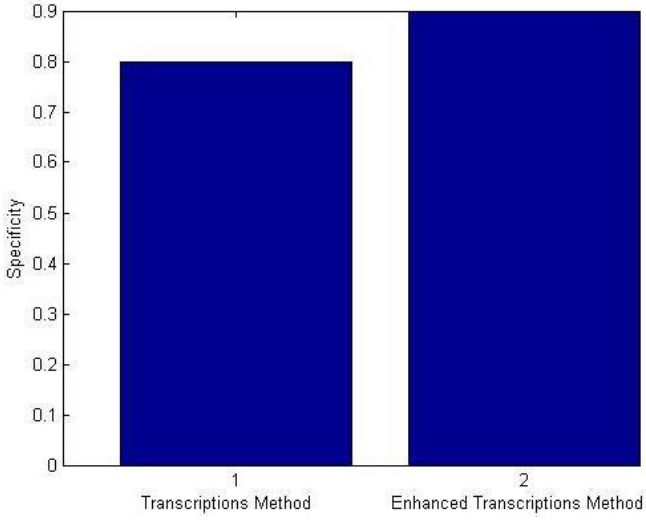

Fig. 9: Specificity

The specificity of the previous (Transcriptions Method) and proposed (Enhanced Transcriptions Method) is shown in figure 9. The accuracy of previous method is 0.79 and proposed method is 0.89.

VI. CONCLUSION AND FUTURE SCOPE

The proposed Enhanced Transcriptions Method is developed using Microsoft Visual Studio in VB.Net Language. Firstly word indexing is performed for the predefined words then corresponding speech signal is detected and errors in words are calculated using Euclidean distance. The results of the proposed work shows that Enhanced Transcriptions Method has more accuracy 89% as compared to previous Transcriptions Method 79%. The value of specificity for proposed method is 0.89 and for previous method is 0.79.

REFERENCES

[1] Faria, A., 2003. Applied phonetics: Portuguese text-to-speech. University of California, Berkeley.
[2] Weerasinghe, R., Wasala, A., Welgama, V. and Gamage, K., 2007, September. Festival-si: A sinhala text-to-speech system. In International Conference on Text, Speech and Dialogue (pp. 472-479). Springer Berlin Heidelberg.
[3] Bali, K., Talukdar, P.P., Krishna, N.S. and Ramakrishnan, A.G., 2004. Tools for the development of a Hindi speech synthesis system. In Fifth ISCA Workshop on Speech Synthesis.
[4] Kayte, S.N., 2015. Festival and Festvox Framework Tools for Marathi Text-to-Speech Synthesis. International Journal of Computer Applications, 132(4), pp.38-43.
[5] Deepa, S.R., Bali, K., Ramakrishnan, A.G. and Talukdar, P.P., 2004. Automatic Generation of Compound Word Lexicon for Hindi Speech Synthesis. In LREC.
[6] Kayte, S.N., 2015. Text To Speech for Marathi Language using Transcriptions Theory. International Journal of Computer Applications, 131(6), pp.39-41.
[7] Panda, S.P., Nayak, A.K. and Patnaik, S., 2015. Text-to-speech synthesis with an Indian language perspective. International Journal of Grid and Utility Computing, 6(3-4), pp.170-178.
[8] En.wikipedia.org. (2017). Speech synthesis. [online] Available at: https://en.wikipedia.org/wiki/Speech_synthesis [Accessed 22 Mar. 2017].
[9] En.wikipedia.org. (2017). International Phonetic Alphabet. [online] Available at: https://en.wikipedia.org/wiki/International_Phonetic_Alphabet [Accessed 22 Mar. 2017].
[10] Search Security. (2017). What is tokenization? - Definition from WhatIs.com. [online] Available at: http://searchsecurity.techtarget.com/definition/tokenization [Accessed 22 Mar. 2017].
[11] En.wikipedia.org. (2017). Transcription (biology). [online] Available at: https://en.wikipedia.org/wiki/Transcription_(biology) [Accessed 22 Mar. 2017].
[12] Barnett, M.P. and Ruhsam, W.M., 1968. A natural language programming system for text processing. IEEE transactions on engineering writing and speech, 11(2), pp.45-52.
[13] Fan, H.T., Hung, J.W., Lu, X., Wang, S.S. and Tsao, Y., 2014, May. Speech enhancement using segmental nonnegative matrix factorization. In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on (pp. 4483-4487). IEEE.

