

# Development of TTS for Marathi Language using Concatenation Approach

Mali Neha<sup>1</sup> Suryawanshi Prajakta<sup>2</sup> Vaishnav Rutuja<sup>3</sup> Prof. S.A. Itkarkar<sup>4</sup>

<sup>1,2,3,4</sup>Department of Electronics and Telecommunication Engineering

<sup>1,2,3,4</sup>BVCOEW, Pune

**Abstract**— Interaction between human and computer goes on increasing day-by-day. Computer can communicate with Humans through voice. This conversation of human and computer is possible through a system which is known as Text-To-Speech system. It is a research field that has received a lot of attention. The Text-To-Speech (TTS) System is the process of converting written text into speech. Text to Speech System is developing technology which has been used in many real time applications like e-learning, for Handicapped, Blind, Illiterate people. The Text-To-Speech systems are commercially available for Global (English) language; one of them is IVONA system. And fewer systems are available for Local (Marathi, Hindi, Kannada etc) languages. Some of them are READSPEAKER, iSPEECH, VACHAK. In this paper we are representing the idea of converting Text into Speech for MARATHI language. This Text-To-Speech system uses a Word based Concatenation synthesis. In which pre-recorded words from the database are concatenated to produce speech using MATLAB. The proposed work has some limitations of number of words. But we can further increase our word database depending on user requirement.

**Key words:** Text-To-Speech (TTS), Concatenation

## I. INTRODUCTION

Now a day's people want much more easy things to reduce their efforts and also time. They want relaxation in day-to-day life. The Text-to Speech (TTS) System is a system through which computer can communicate with human being. This system is becoming more and more popular as the interaction between human and computer goes on increasing. It involves one of the best Options found out till date is the ability of a computer to speak to Humans. How computer can speak to humans? This is precisely the question we explore in this paper. Text-To-Speech is a process through which input text is analyzed, processed and understood then the text is rendered as an audio, or it can be a procedure performed by a computer to estimate some kind of a presentation of the speech signal for a given text input.

Marathi is an Indo-Aryan language spoken predominantly by Marathi people of Maharashtra. It is the official and co-official language in Maharashtra and Goa states of Western India respectively. Marathi Language is spoken by more than 73 million people across the world. As there are fewer systems available for Marathi Text-to-Speech. So here we are proposing a Text-to-Speech system for Marathi language in human voice.

During the learning process at all school levels, it is used as an electronic learning (e-learning) or distance learning. The special importance is for Handicapped people because it can soften numerous difficulties they face. People with vision disabilities or totally blind people, have many limitations and restrictions in meeting their normal daily activities if not assisted by their relatives or their caretakers. Also, in the education process, such people are unable to use

ordinary books or various video records which are related to the process of their Education or Entertainment.

### A. Types of Implementation:

The process of transforming text into speech contains coarsely two phases: first the text goes through analysis and then the resulting information is used to generate the speech signal. The following subsections describe the main principles of the three most commonly used speech synthesis methods: formant synthesis, concatenative synthesis, and articulatory synthesis.

#### 1) Formant Synthesis:

This is the oldest method for speech synthesis, and it dominated the synthesis implementations for a long time. In this speech production mechanism are mainly based on the source filter theory. In these models, a linear filter simulates the vocal tract, which is driven by a source. This approach uses several formant filters to filter the excitation signals from a variable frequency signal source (for voice speech components) and a white noise source (for unvoiced speech). This method is sometimes called rules-based synthesis.

#### 2) Concatenative Synthesis:

Concatenative synthesis basically means the concatenation or stringing together of the segments of already recorded speech. Generally, concatenative synthesis can generate natural-sounding synthesized speech. The system records the speech and extracts the segments by an automatic segmentation process. These segments are joined together resulting in audible glitches in the output. Concatenative synthesis further classified as given below:

- 1) Unit selection synthesis
- 2) Diphone synthesis
- 3) Domain-specific synthesis
- 4) Phoneme-based synthesis

#### 3) Articulatory Synthesis:

Articulatory synthesis refers to the techniques for synthesizing speech based on models of the human vocal tract and the organs taking part in the articulation processes. The techniques discussed so far attempt to generate a signal that is perceptually optimal. They produce a wave whose spectrum is as close as possible to the real speech signal without giving consideration to simulating any aspect of the human vocal tract. Such techniques attempt to limit the negative effects of the coarticulation.

Out of these techniques by using "Concatenative Domain-specific Synthesis" proposed system is implement. In which prerecorded words are concatenated to create complete utterances.

## II. STRATEGY FOLLOWED

The proposed system presents an implementation of text to Speech for Marathi speech signal using concatenation approach. Firstly by the use of microphone words are recorded. The recorded sound saved in .wav format. User can

manually enter text in Marathi language using English character. Audio file will be generated as an output which represents speech equivalent to given text. These processes are done with the help of MATLAB.

**A. Algorithm:**

**1) For Database Creation:**

- 1) Record the words using microphone.
- 2) Save recorded words in computer using .wav format.
- 3) Process the database to remove silence.
- 4) Save the processed audio database in computer.

**B. For Text to Speech:**

- 1) Give the text input to the system using keyboard
- 2) Entered text will be searched in audio database.
- 3) Ones the corresponding audio file of input text found in database it will concatenate all the audio files .
- 4) Play concatenated .wav files to generate the speech equivalent to input text.

**III. BLOCK DIAGRAM**

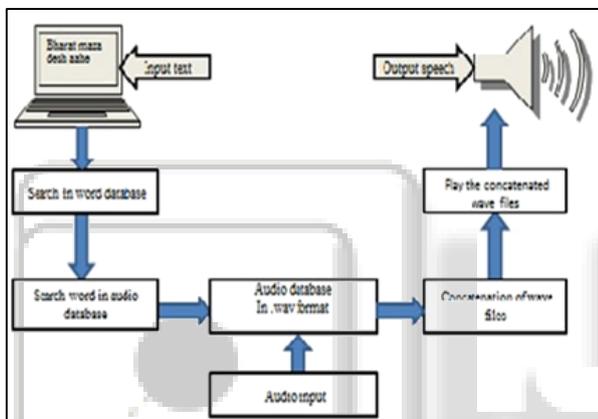


Fig. 1: Block diagram of proposed system

**A. Creation of Database:**

This is the very basic step of proposed system as it is prerecorded type system. We have to record the words. There should be some precautions to be take while recording

- 1) Sound recording should be done in noiseless chamber to avoid noise interference.
- 2) Use good quality microphone to record speech.
- 3) Recorded speech saved in .wav format.

**B. Database Processing:**

Ones the recording done we have to remove unwanted content from the recorded word, like silence. The memory required for words after recording and after recording word processing is as given in below table:

WORDS	RECORDED SIZE	AFTER PROCESSING
Kharedi	64.6kbps	11.3kbps
Nahi	43.1kbps	9.6kbps
Aani	43.1kbps	8.41kbps

Table 1: Size variations in words

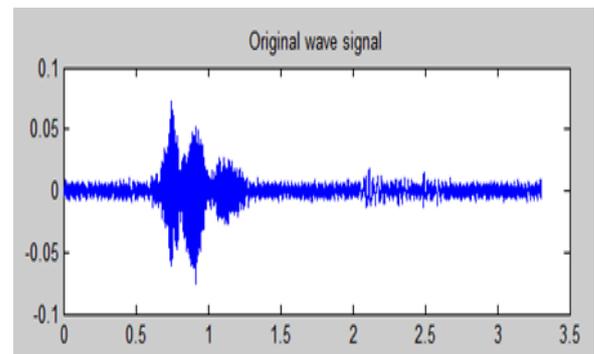


Fig. 2: Waveform before processing (kharedi)

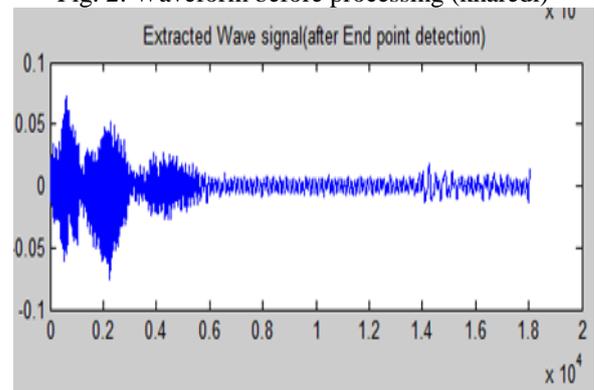


Fig. 3: Waveform after processing (kharedi)

**C. Input to The System:**

Input to the proposed system is given by using keyboard. Text entered in text file is in Marathi language using English characters.

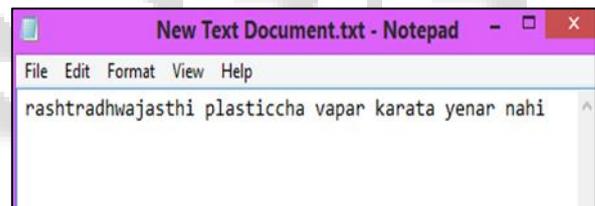


Fig. 4: System Input

The spelling of entered word should be match with the spelling of that word stored in database as there is no provision is provided for spell checking.

**D. Searching and Concatenation:**

Text entered in text file will be searched in a recorded database. Ones the correct match found we go for concatenation. Concatenation is the operation of joining sound waves end-to-end. In concatenation method, the synthetic speech is made by joining a speaker's natural short term waveform segments which have accumulated in a large speech corpus (database) in advanced. In a concatenation synthesis approach, the TTS module obtains the sound files from an acoustic inventory corresponding to the string of words and concatenates them.

**IV. CONCLUSION**

As per the goal of proposed system is an attempt made to show how the computer speaks out the Marathi. Here the user can give input text by using keyboard and he can listen to his text. By developing such systems, relationship between human and computer becomes much closer.

REFERENCES

- [1] 1.Surendra P. Ramteke, 2.Gunjal Oza, 3.Nilima P. Patil, "Development of TTS for Marathi Speech Signal Based on Prosody and Concatenation Approach".International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 10, December- 2012
- [2] 1 Tapas Kumar Patra ,2 Biplab Patra ,3 Puspanjali Mohapatra, "Text to Speech Conversion with Phonematic Concatenation" International Journal of Electronics Communication and Computer Technology (IJECCCT) Volume 2 Issue 5 (September 2012)
- [3] V.Ramu Reddy, K.Sreenivasa Rao , "Prosody modeling for syllable based text-to-speech synthesis" ,24 July 2015 , at ScienceDirect .
- [4] Prof .Mrs. Smita Kawachale1, Prof Dr.J. S. Chitode2, "Relative Functional Comparison of Neural and Non-Neural Approaches for Syllable Segmentation in Devnagari TTS System", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012 .
- [5] Yannis Stylianou, Member, IEEE "Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis." IEEE transactions on speech and audio processing, vol. 9, no. 1, january 2001
- [6] Mr.S.D.Shirbahadurkar, Dr.D.S.Bormane. "Speech Synthesizer Using Concatenative Synthesis Strategy for Marathi language (Spoken in Maharashtra, India)"Int. J. of Recent Trends in Engineering and Technology, Vol. 2, No. 4, Nov 2009
- [7] Darshna Badhe, P. M. Ghate. "Marathi Text to Speech Synthesis – Using Matlab®" IJCSN International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015
- [8] Technology Development For Indian Languages, Government of India. [http://tdil.mit.gov.in/publication/vishvabharat]
- [9] Marathi language – Wikipedia, the free encyclopedia. [https://en.m.wikipedia.org]