



TEXT-TO-SPEECH SYSTEM FOR GUJARATI LANGUAGE

¹Jashin Marina Varghese, ²Smita Hande

Electronics and Telecommunication

Fr. Conceicao Rodrigues Institute of Technology

Vashi, India

Email: ¹Jashinvarghese@gmail.com, ²handesmita@yahoo.com

Abstract— In this paper we describe Speech synthesis which also called as Text – to-speech synthesis. The speech synthesized is the artificial generation of human speech from text (Gujarati text). The required database in the conversion of character-to-sound is recorded by PRAAT software and is saved in .wav format in the directory. The smallest segment of the recorded voice is concatenated to produce speech of the desired words. The main aim of using this method is, it is simple. Concatenation technique is used and software used is MATLAB.

Keywords—Concatenative synthesis, gujarati, praat, tts, unicode, PESQ.

I. INTRODUCTION

The word ‘Synthesis’ is defined by the Dictionary as ‘the putting together of parts or elements so as to form a whole’. Speech synthesis generally refers to the artificial generation of human voice – either in the form of speech or in other forms such as a song. The computer system used for speech synthesis is known as a speech synthesizer. There are several types of speech synthesizers (both hardware based and software based) with different underlying technologies. For example, a TTS (Text to Speech) system converts normal language text into human speech, while there are other systems that can convert phonetic transcriptions into speech. The basic principle behind speech synthesis is known as the source-filter theory of speech production – that is, the information about the voice source and

the vocal tract filter can be separated from each other.

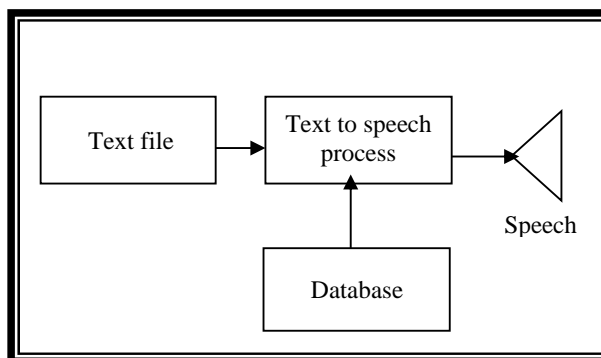


Fig. 1. Block diagram of normal text-to-speech system.

Today, speech synthesizers are a common feature in most operating systems. Speech synthesis applications have also made computing related services more inclusive by allowing access to people with visual impairments or reading disabilities.

The quality of a speech synthesizer is measured based on two primary factors – its similarity to normal human speech (naturalness) and its intelligibility (ease of understanding by the listener). Ideally, a speech synthesizer should be both natural and intelligible, and speech synthesis systems always attempt to maximize both characteristics. Evaluation of speech is done through MOS mainly, but in this paper evaluation of speech is planned with PESQ method.

II. OVERVIEW OF GUJARATI LANGUAGE

Gujarati (ગુજરાતી) is an Indo-Aryan language native to the West Indian region of Gujarat. The Gujarati script was adapted from the Devanagari script to write the Gujarati language. It is part of the greater Indo-European language family. Gujarati is descended from Old Gujarati. Gujarati is a modern IA (Indo-Aryan) language evolved from Sanskrit.

In Gujarati language there are 33 consonants and 13 vowels. The phonemes are divided into two type's vowels (swaras) and consonants (vyanjanas).

vowels (swaras): Vowels are the independently existing letters. Vowels sound cannot be modified.

V VF > . p μ k VV{ VM VF{ V \ Vo

Consonants (vyanjanas): Consonants are those which are depend on vowels to form their independent letters. The sound of the consonant can be modified by combining vowels with them.

SBU3RKHh 890-6TYN WG50AEDI
Z,JXQF;C/ 1F7

III. SYNTHESIZER TECHNOLOGY USED

A. Concatenative Synthesis

Concatenation is also called as cut and paste synthesis in which short segment of speech are selected from a pre-recorded database and joined one after another to produce the desired utterance. In this the quality of speech is high but there many b some limitation due to the memory capacity. The longer selected units are the fewer problematic concatenation point will occur in the synthesis speech, but at the same time memory requirement increases. The length of unit selections can be phonemes and diaphones because they are short enough to attain sufficient flexibility and to keep the memory requirement reasonable. Their use of provide good quality speech to take account of co-articulation cause they contain third transition from one phonemes to another and later half of the later phonemes. The use of phonemes is the most flexible way of generating various utterances.[6]

In phonemes and diaphone concatenation is the greatest challenge in the continuity and to avoid audible distortion caused by the difference between successive segments, at

least the fundament frequency and the intensity of the segment must b controlled. It is a troublesome process of creating the database from which unit be selected. [2][3]

IV. CREATION OF DATABASE

For Gujarati language there are 426 phonemes including vowels and consonant and their combination. The combination include here is C, CV, V, VC where C is the consonant and V is the vowel. The algorithm used to merge the characters or word is concatenation method. The entire 426 phonemes database is created in the form of wave files (.wav) by using PRAAT software and Fs is 44100 Hz Where 13 phonemes are vowels and 34 phonemes are consonant and rest are their combinations.

TABLE 1. Phonemes along with the letters with examples (Gujarati)

Sr. no	Phonemes	Example
1.	SF	SFSF
2.	9	9I/IM
3.	,L	S,L

TABLE 2. Examples of all possible combinations of 'S'.

Sr. no	Phonemes combinations
1.	S + V = S
2.	S + VF = SF
3.	S + > = IS
4.	S + . = SL
5.	S + p = S]
6.	S + μ = S)
7.	S + k = S=
8.	S + V[= S[
9.	S + V{ = S{
10.	S + VM = SM
11.	S + VF{ = SF{
12.	S + V\ = S\

A. FOR GUJARATI LANGUAGE

The process of transformation from text to speech contains mainly two phrases.

- The text analysis,

- Then the resulting information is used to generate the speech signal (speech synthesis).

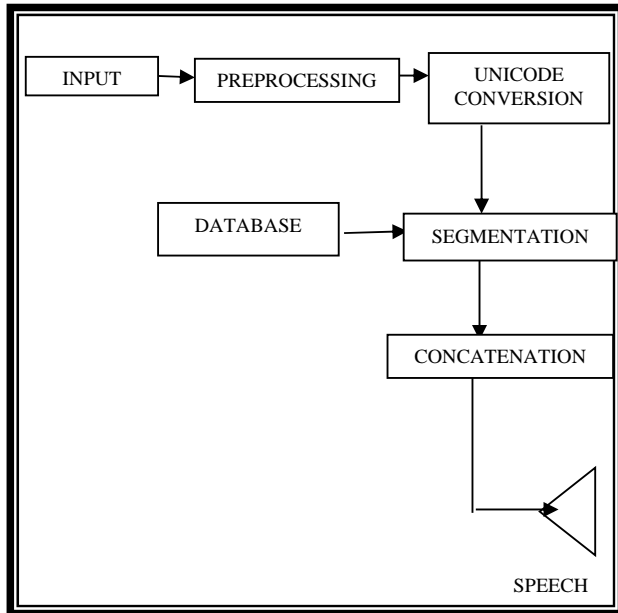


Fig. 2. Block diagram for TTS for Gujarati language.

Text analysis means syntactic and semantic analysis is done and to identify the font in which the given text is encoded and applying the Unicode to given font is done in Unicode conversion and the phonetic analysis is done in this which the graphemes are converted into phonemes. Then prosodic analysis in which pitch and duration attachment is occurred. Then finally speech synthesis in voice rendering is done and that sound is produced. The pre-recorded phonemes are concatenated according to the given text characters.

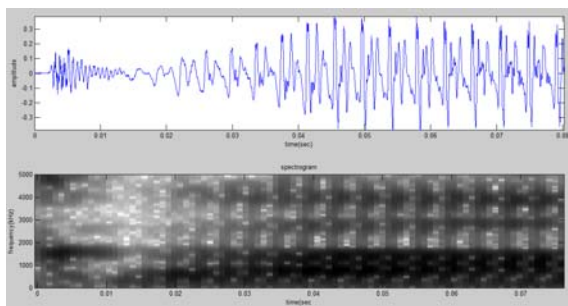


Fig. 3. Output of speech "S"

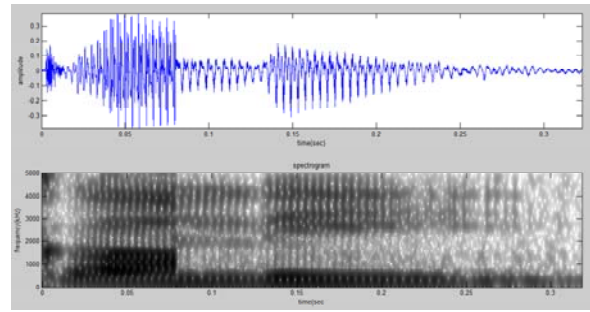


Fig. 4. Output of speech "S,L"

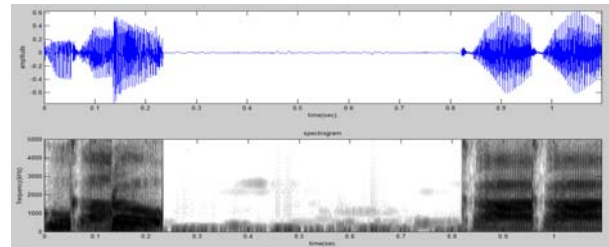


Fig. 5. Output of speech "SJSOM SFSF".

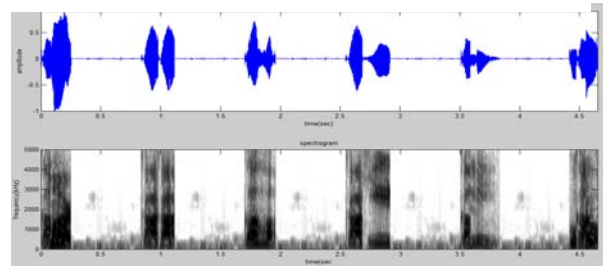


Fig. 6. Output of speech "SD/ SFSF So SFSL S,L SJSOM".

VI. QUALITY TEST

To measure the speech quality there are two approaches: subjective and objective. The subjective listening tests a subject hears a recorded speech of a processed voice file and the quality using an opinion scale and the MOS is then calculated as an average of participant scores. Objective models have been developed to provide machine-based automatic assessment of the speech quality score. The measure is derived from the input text and the natural speech corpus and is inversely proportional to the overall speech quality-the higher number of concatenations, the lower is the quality. Alternatively, signals based measures have been proposed which focus on the computing spectral distances between the target synthesized speech signal and its original natural speech.

Since perceptual evaluation of speech quality(PESQ) is a more popular tool and has been widely deployed in the industry and many

research area and is composed of five scores of objective quality.

TABLE 3. Score of objective listening test

rating	Description
5	Very natural, no degradation
4	Fairly natural , little degradation
3	Somewhat natural, somewhat degraded
2	Fairly unnatural, fairly degraded
1	Fully unnatural, fully degraded

TABLE 4. Objective listening test results.

Test set	PESQ
SFSF	4.14
So	4.35
SD/	3.36
9I/IM	3.914
RS,L	4.681
CS	3.677

VII. CONCLUSION

This paper discusses the design and development of Gujarati text to speech for concatenation based TTS system. This paper have selected C,CV,V,VC patterns only. Since this method is very easy and efficient to implement as compared with other methods. And can conclude that the speech produced is preserving naturalness and good quality based on the objective quality test results. This can be applied for different languages with more naturalists in them. As future work can b apply on word documents, and with advanced version can apply on PDF, scanned data etc. by some modifications.

Acknowledgment

Author would like to express gratitude to N. Oza for spearing her precious time for the recording of the database in Gujarati language.

References

[1] S. Lemmetty S, "Review of Speech Synthesis Technology", M.S. Thesis, Dept Elec.Helsinki University of Technology, 1999.

[2] N.Swetha and k.Anuradha: "Text-to-speech conversion" International journal of advanced trends in compuer science and engineering, Vol.2 , No.6, pp : 269-278 (2013).

[3] A.Chauhan, V.Chauhan, G.Singh, C.Coudhary and P. Arya: "Design and development of a text-to-speech synthesizer system" International journal of electronics and communication technology, Vol. 2, Issue 3,pp 322-326, Sept. 2011.

[4] A.Indumathi and Dr.E.Chandra."Survey on speech synthesis" signal processing:An

internationaljournal(SPIJ),volume(6):issue(5):p p 140-145,2012.

[5] M.Ahmed and S.nisar." text to speech synthesis using phoneme concatenation" International journal of electronics and communication technology, Vol. 2, Issue 3,pp 193-197, feb 2014

[6] S.D.Suryawanshi, R.R.Itkarkar and D.T.Mane: "High quality text to speech synthesizer using phonetic integration" International Journal of Advanced Research In Electronics And Communication Engineering. Volume 3, Issue 2,February 2014, pp 77-82.

[7] D.sasirekha and E.chandra: "Text-to-speech : A simple tutorial" International journal of advanced trends in compuer science and engineering, Vol.2 , Issue-1, pp : 269-278 ,March 2012.

[8] S.Ramteke And G.Oza" Development Of Tts For Marathi Speech Based Prosody And Concatenation Approach" " International Journal Of Electronics And Communication Technology, Vol. 1, Issue 10,pp 1-6, Dec. 2012.

[9] R.Mengko and A. Ayuningtyas" Indonesian TTS system using syllable concatenation: Speech Optimization"IEEE Transaction On Instrumentation, Communication and Biomedical Engineering, November 2013, pp 412-415.

[10] A.Jocab And P.Mythili:" Developing a Child Friendly Text-To-Speech System" Hindwani Publishing Corporation Advanved in Human-computer Iteration. Volume 2008, Article ID 597971,pp 6

[11] S.Ramteke And G.Oza" Development Of TTS For Marathi Speech Based Prosody And Concatenation Approach" " International Journal Of Electronics And Communication Technology, Vol. 1, Issue 10,pp 1-6, Dec. 2012.

[12] T.Shreekanth, V. Udayashankara And C. Arun Kumar. "A Unit Slelection Based Text To Speech Synthesis System Using Syllable as a Basic Unit" IOSR Journal Of VLSI And Signal Processing (IOSR-JVSP), Volume 4,Issue 4,Jul-Aug 2014,pp 49-57.

[13] M. Cernak and M. Rusko "An Evaluation of Speech Using PESQ Measures" Forum Acusticum 2005 Budapest.

[14]S. Moller and T. H. Falk "Quality Prediction for Synthesized Speech: Comparison of Approaches" NAG/DAGA 2009-Rotterdam.Pp1168-1171.